



Ardian - 2106638173
Data Mining
Assignment 4 - Association Rule Mining

Diberikan suatu data transaksi penjualan yang berhasil dilakukan pada suatu toko elektronik sebagai berikut:

- Transaction 1: {Laptop, Mouse, Keyboard, Headphones}
 - Transaction 2: {Laptop, Mouse, Keyboard}
 - Transaction 3: {Mouse, Keyboard, Headphones}
 - Transaction 4: {Laptop, Keyboard, Headphones}
 - Transaction 5: {Laptop, Mouse, Keyboard, Webcam}
1. Dengan minimum support = 0.6, jalankan algoritma secara naif untuk mendapatkan frequent itemset dengan memanfaatkan data transaksi diatas! Jelaskan setiap langkah yang perlu dilakukan.

Jawaban:

Pertama, berdasarkan Apriori Principle, kita bisa mengfilter terlebih dahulu item yang memiliki support < minimum support, yaitu item yang transaksi-nya kurang dari 3 kali. Dengan begitu, kita tersisa dengan Laptop, Keyboard, Mouse, dan Headphones. Kemudian kita hitung nilai support untuk setiap kemungkinan itemset. Berikut hasil perhitungannya:

Laptop	Keyboard	Mouse	Headphones	n Transaction	Support
0	0	0	0	0	-
0	0	0	1	3	0.6
0	0	1	0	4	0.8
0	0	1	1	2	0.4
0	1	0	0	5	1
0	1	0	1	3	0.6
0	1	1	0	4	0.8
0	1	1	1	2	0.4
1	0	0	0	4	0.8
1	0	0	1	2	0.4
1	0	1	0	3	0.6
1	0	1	1	1	0.2
1	1	0	0	4	0.8

1	1	0	1	2	0.4
1	1	1	0	3	0.6
1	1	1	1	1	0.2

Dari hasil perhitungan di atas, dengan minimum support 0.6, kita mendapatkan 3 frequent itemsets, yaitu:

1. {Headphones}
 2. {Mouse}
 3. {Keyboard}
 4. {Keyboard, Headphones}
 5. {Keyboard, Mouse}
 6. {Laptop}
 7. {Laptop, Mouse}
 8. {Laptop, Keyboard}
 9. {Laptop, Keyboard, Mouse}
2. Dengan minimum support = 0.6, jalankan algoritma Apriori dengan menuliskan F sebagai frequent itemsets pada suatu K dan C sebagai kandidat untuk frequent itemsets selanjutnya dengan K adalah jumlah item dalam himpunan berdasarkan data transaksi diatas!

Jawaban:

Minimum support adalah 0.6, yang berarti syarat frequent itemset adalah $\text{count} \geq$

3. Dan berdasarkan prinsip apriori, saya akan langsung mengabaikan superset yang super subsetnya tidak memenuhi minimum support. Maka:

1 itemset	count	2-itemset	count	3-itemset	count	4-itemset	count
{Laptop}	4	{Laptop, Keyboard}	4	{Laptop, Keyboard, Mouse}	3		
{Keyboard}	5	{Laptop, Mouse}	3				
{Mouse}	4	{Laptop, Headphones}	2				
{Headphones}	3	{Keyboard, Mouse}	4				
{Webcam}	1	{Keyboard, Headphones}	3				
		{Mouse, Headphones}	2				

⇒ **Frequent itemsets:**

1. {Laptop} dengan support 0.8
2. {Keyboard} dengan support 1
3. {Mouse} dengan support 0.8

4. {Headphones} dengan support 0.6
 5. {Laptop, Keyboard} dengan support 0.8
 6. {Laptop, Mouse} dengan support 0.6
 7. {Keyboard, Mouse} dengan support 0.8
 8. {Keyboard, Headphones} dengan support 0.6
 9. {Laptop, Keyboard, Mouse} dengan support 0.6
3. Dengan minimum confidence sebesar 0.5, temukan semua aturan asosiasi dari frequent sets yang telah ditemukan sebelumnya!

Jawaban:

Berikut perhitungan confidence untuk setiap rule candidates:

1. $\text{conf}(\{\text{Laptop}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard}\})}{\text{sup}(\{\text{Laptop}\})} = 0.8/0.8 = 1$
2. $\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Laptop}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.8/1 = 0.8$
3. $\text{conf}(\{\text{Laptop}\} \Rightarrow \{\text{Mouse}\}) = \frac{\text{sup}(\{\text{Laptop, Mouse}\})}{\text{sup}(\{\text{Laptop}\})} = 0.6/0.8 = 0.75$
4. $\text{conf}(\{\text{Mouse}\} \Rightarrow \{\text{Laptop}\}) = \frac{\text{sup}(\{\text{Laptop, Mouse}\})}{\text{sup}(\{\text{Mouse}\})} = 0.6/0.8 = 0.75$
5. $\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Mouse}\}) = \frac{\text{sup}(\{\text{Keyboard, Mouse}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.8/1 = 0.8$
6. $\text{conf}(\{\text{Mouse}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Keyboard, Mouse}\})}{\text{sup}(\{\text{Mouse}\})} = 0.8/0.8 = 1$
7. $\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Headphones}\}) = \frac{\text{sup}(\{\text{Keyboard, Headphones}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.6/1 = 0.6$
8. $\text{conf}(\{\text{Headphones}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Keyboard, Headphones}\})}{\text{sup}(\{\text{Headphones}\})} = 0.6/0.6 = 1$
9. $\text{conf}(\{\text{Laptop, Keyboard}\} \Rightarrow \{\text{Mouse}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard, Mouse}\})}{\text{sup}(\{\text{Laptop, Keyboard}\})} = 0.6/0.8 = 0.75$
10. $\text{conf}(\{\text{Mouse}\} \Rightarrow \{\text{Laptop, Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard, Mouse}\})}{\text{sup}(\{\text{Mouse}\})} = 0.6/0.8 = 0.75$
11. $\text{conf}(\{\text{Laptop, Mouse}\} \Rightarrow \{\text{Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard, Mouse}\})}{\text{sup}(\{\text{Laptop, Mouse}\})} = 0.6/0.6 = 1$
12. $\text{conf}(\{\text{Keyboard}\} \Rightarrow \{\text{Laptop, Mouse}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard, Mouse}\})}{\text{sup}(\{\text{Keyboard}\})} = 0.6/1 = 0.6$
13. $\text{conf}(\{\text{Mouse, Keyboard}\} \Rightarrow \{\text{Laptop}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard, Mouse}\})}{\text{sup}(\{\text{Mouse, Keyboard}\})} = 0.6/0.8 = 0.75$
14. $\text{conf}(\{\text{Laptop}\} \Rightarrow \{\text{Mouse, Keyboard}\}) = \frac{\text{sup}(\{\text{Laptop, Keyboard, Mouse}\})}{\text{sup}(\{\text{Laptop}\})} = 0.6/0.8 = 0.75$

⇒ Karena semua kandidat memenuhi syarat confidence, maka semua kandidat merupakan strong association rule.

4. Untuk data transaksi yang lebih besar, pengerjaan dengan tangan akan sangat tidak efisien sehingga diperlukan untuk dapat mengimplementasikan algoritma untuk

menemukan frequent itemsets, implementasikan algoritma untuk menghitung frequent itemsets tersebut tanpa menggunakan library module (make it from scratch)!

Jawaban:

Berikut link notebook full implementation-nya:

https://colab.research.google.com/drive/1xj6Ay5w-tEHSTVfAY8Ds33TTEFEbqDzF?usp=share_link

Snippet main functionnya:

```
def Apriori(transactions_df, minsup):
    frequent_item_sets_per_level = defaultdict(list)

    # Extract item names and create a dictionary to map items to indices
    item_list = list(transactions_df.columns)
    item_dict = {item: i + 1 for i, item in enumerate(item_list)}

    # Process transactions dataframe to be processable
    transactions = []
    for _, row in transactions_df.iterrows():
        transaction = set()
        for item in item_dict:
            if row[item] == 1:
                transaction.add(item)
        transactions.append(transaction)

    # Generate frequent item sets for level 1
    for item in item_list:
        support = get_support(transactions, {item})
        if support >= minsup:
            frequent_item_sets_per_level[1].append((item, support))

    # Generate frequent item sets for higher levels
    for level in range(2, len(item_list) + 1):
        current_level_candidates = self_join(frequent_item_sets_per_level, level)
        post_pruning_candidates = prune(frequent_item_sets_per_level, level, current_level_candidates)
        if len(post_pruning_candidates) == 0:
            break

        for item_set in post_pruning_candidates:
            support = get_support(transactions, item_set)
            if support >= minsup:
                frequent_item_sets_per_level[level].append((item_set, support))

    return frequent_item_sets_per_level
```

Hasil test algoritma menggunakan data transaksi yang diberikan dengan minimum support 0.6:

```
Level 1 frequent itemsets:
1. {Laptop} with 0.8 support
2. {Mouse} with 0.8 support
3. {Keyboard} with 1.0 support
4. {Headphones} with 0.6 support

Level 2 frequent itemsets:
1. {Mouse, Laptop} with 0.6 support
2. {Laptop, Keyboard} with 0.8 support
3. {Mouse, Keyboard} with 0.8 support
4. {Keyboard, Headphones} with 0.6 support

Level 3 frequent itemsets:
1. {Mouse, Laptop, Keyboard} with 0.6 support
```

5. Menurut Anda, apa kelebihan dan kekurangan dari pemanfaatan aturan asosiasi sebagai sistem rekomendasi jika dibandingkan dengan kebanyakan machine learning based algorithm untuk sistem rekomendasi e.g matrix factorization algorithm?

Jawaban:

Kelebihan association rule antara lain adalah simple dan mudah diimplementasikan, efisien, dan sangat interpretable. Dengan rules yang mudah dipahami, pengguna dapat dengan cepat memahami rekomendasi yang diberikan dan memahami alasan dibalik rekomendasi tersebut. Namun, aturan asosiasi memiliki keterbatasan dalam menangani data yang besar dan kompleks.. Di sisi lain, machine learning model memiliki keunggulan dalam menangani data yang besar dan kompleks, serta dalam menangkap hubungan yang kompleks antara item-item. Namun, umumnya model dan rekomendasi yang dihasilkan cukup sulit untuk diinterpretasikan. Selain itu, machine learning memerlukan jumlah data pelatihan yang substansial untuk menghasilkan model yang akurat, yang juga dapat memerlukan sumber daya komputasi yang signifikan. Jadi, aturan asosiasi mungkin lebih cocok untuk kasus-kasus di mana data tidak terlalu kompleks dan juga interpretasi dan kejelasan rekomendasi menjadi prioritas, sementara machine learning lebih sesuai untuk kasus-kasus di mana robustness rekomendasi menjadi prioritas utama.