

用于异构实体消解的层次匹配网络

Hierarchical Matching Network for Heterogeneous Entity Resolution. In: IJCAI 2020.

付成, 韩先培, 孙乐

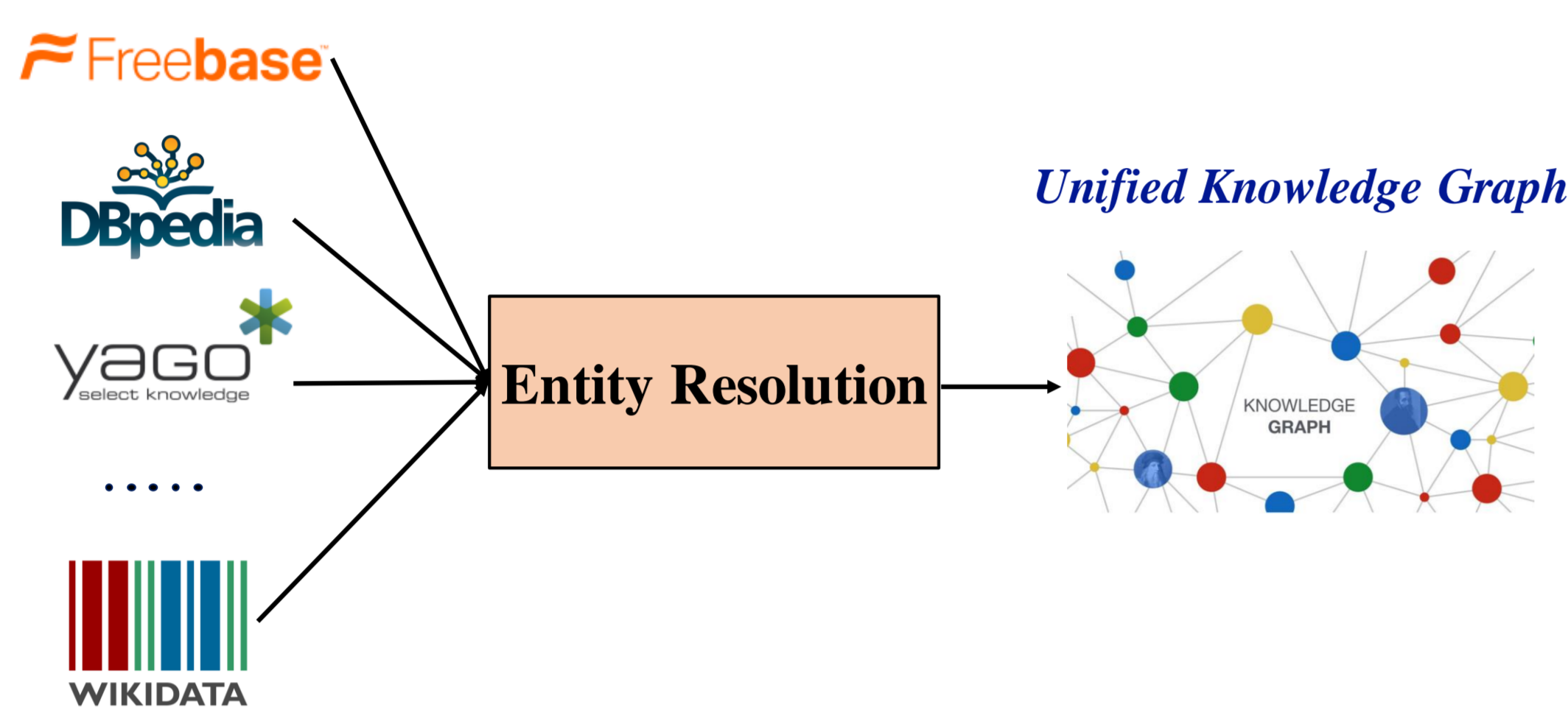
联系人: 韩先培 (xianpei@iscas.ac.cn)

实体消解

➤ 识别不同数据源中指向相同对象的实体

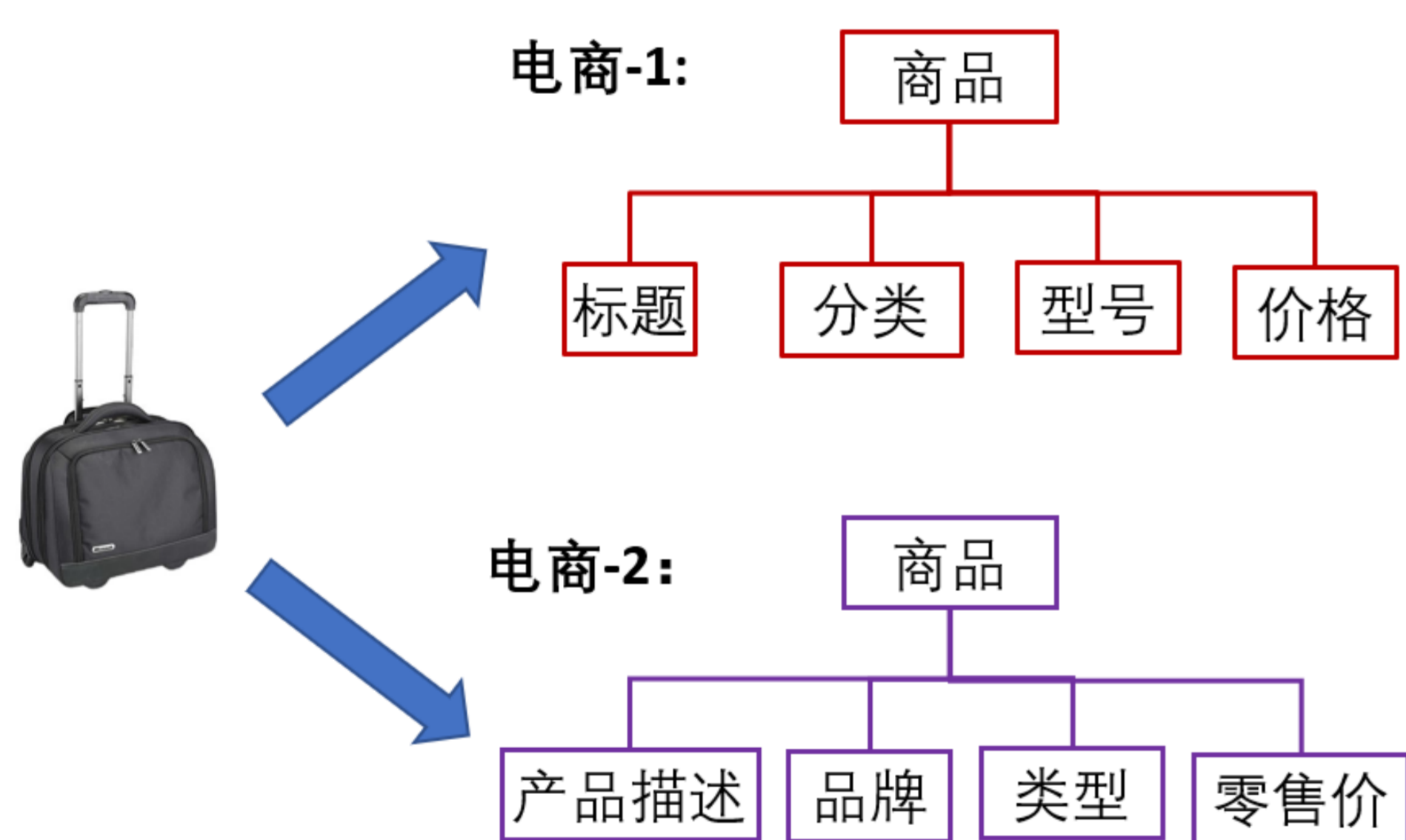
| 姓名 | 职位 | 地址 | | 姓名 | 职位 | 地址 |
|---------------|-------------|--------------|---|---------------|------------|----------------------|
| John Dent | Accountant | Berkeley, CA | ↔ | John | Accountant | Berkeley, California |
| Bush | Salesperson | Phoenix, AZ | ↔ | J. Bush | Marketer | Phoenix, Arizona |
| Frank | Designer | Houston, TX | ✗ | Frank | Designer | Chicago, Illinois |
| Mark A. White | Engineer | Seattle, WA | ✗ | Mark B. White | Engineer | Seattle, Washington |

➤ 桥接和整合不同知识图谱的关键技术

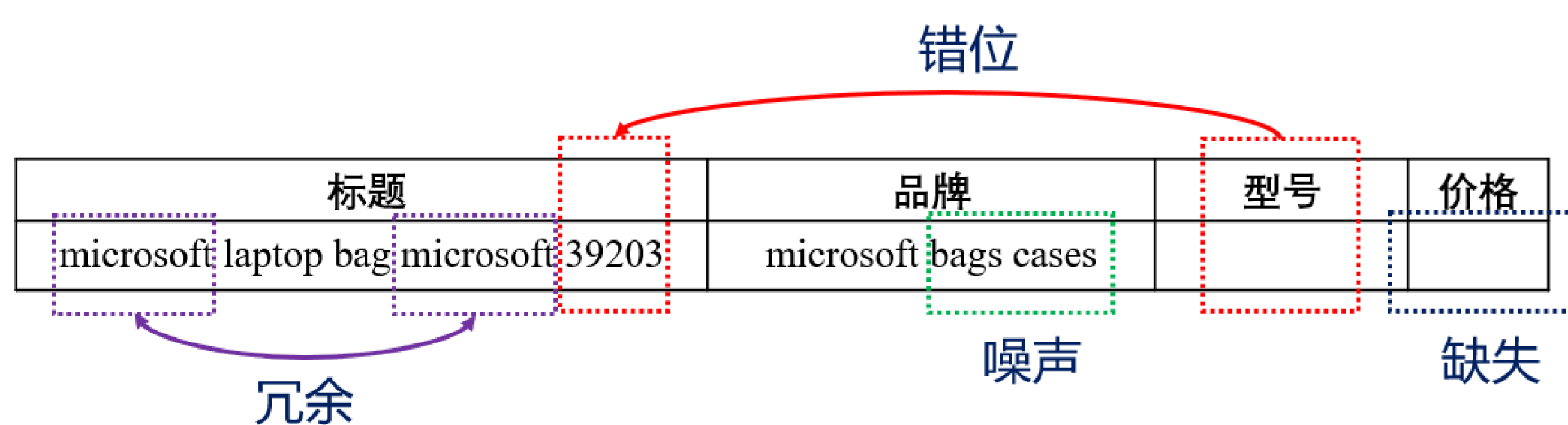


挑战

➤ **模式异构**: 使用不同属性描述相同实体



➤ **脏数据**: 冗余、噪声、缺失、错位...



实验结果

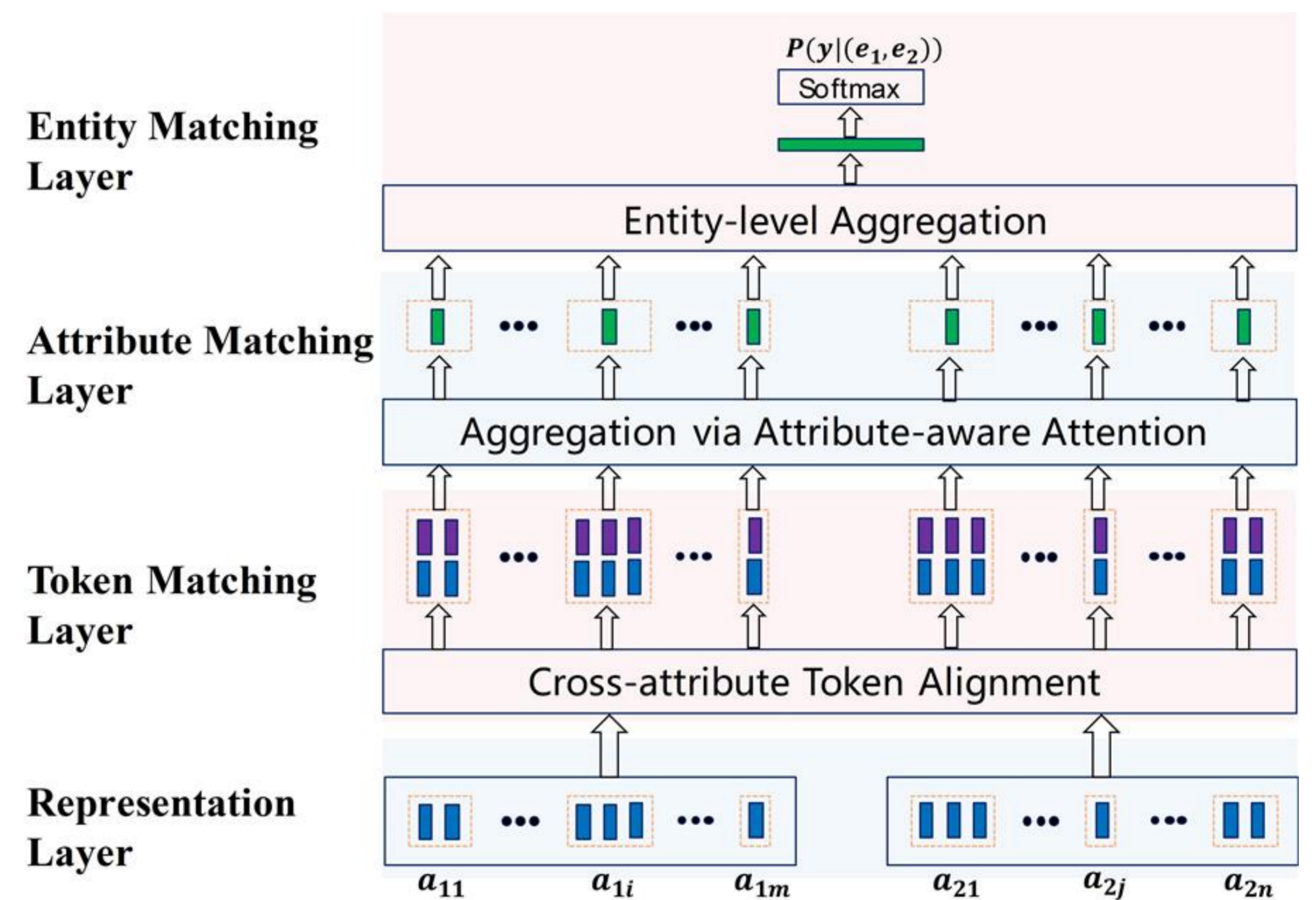
➤ "HierMatcher" 在多种类型的数据集 (同构、异构、脏数据) 上取了当前最好的结果

| Type | Dataset | F1 score | | | | | $\Delta F1$ | $\Delta F1'$ | |
|---------------|-----------------------------------|----------|--------------|------|-----------------|-----------------|-------------|--------------|-------------|
| | | Magellan | Deep-Matcher | MPM | Seq2Seq-Matcher | HierMatcher-ave | | | HierMatcher |
| Homogeneous | Walmart-Amazon ₁ | 71.9 | 67.6 | 73.6 | 78.2 | 77.1 | 81.6 | +8.0 | +3.4 |
| | Amazon-Google | 49.1 | 69.3 | 70.7 | 61.2 | 70.0 | 74.9 | +4.2 | +13.7 |
| | DBLP-ACM ₁ | 98.4 | 98.4 | - | 98.9 | 98.1 | 98.8 | +0.4 | -0.1 |
| | DBLP-Scholar ₁ | 92.3 | 94.7 | - | 95.3 | 94.9 | 95.3 | +0.6 | +0.0 |
| | Average | - | - | - | - | - | - | +3.3 | +4.3 |
| Dirty | Walmart-Amazon ₂ | 37.4 | 53.8 | - | 68.3 | 61.4 | 68.5 | +14.7 | +0.2 |
| | DBLP-ACM ₂ | 91.9 | 98.1 | - | 98.4 | 97.3 | 98.1 | +0.0 | -0.3 |
| | DBLP-Scholar ₂ | 82.5 | 93.8 | - | 94.1 | 93.9 | 94.5 | +0.7 | +0.4 |
| | Average | - | - | - | - | - | - | +5.1 | +0.1 |
| Heterogeneous | Walmart-Amazon ₃ (1-n) | - | 67.1 | - | 75.6 | 67.5 | 80.7 | +13.6 | +5.1 |
| | Walmart-Amazon ₄ (n-n) | - | 63.4 | - | 74.7 | 67.6 | 81.4 | +18.0 | +6.7 |
| | Walmart-Amazon ₅ (n-n) | - | 66.5 | - | 74.4 | 69.0 | 81.0 | +14.5 | +6.6 |
| | Average | - | - | - | - | - | - | +15.4 | +6.1 |

层次匹配网络框架 (HierMatcher)

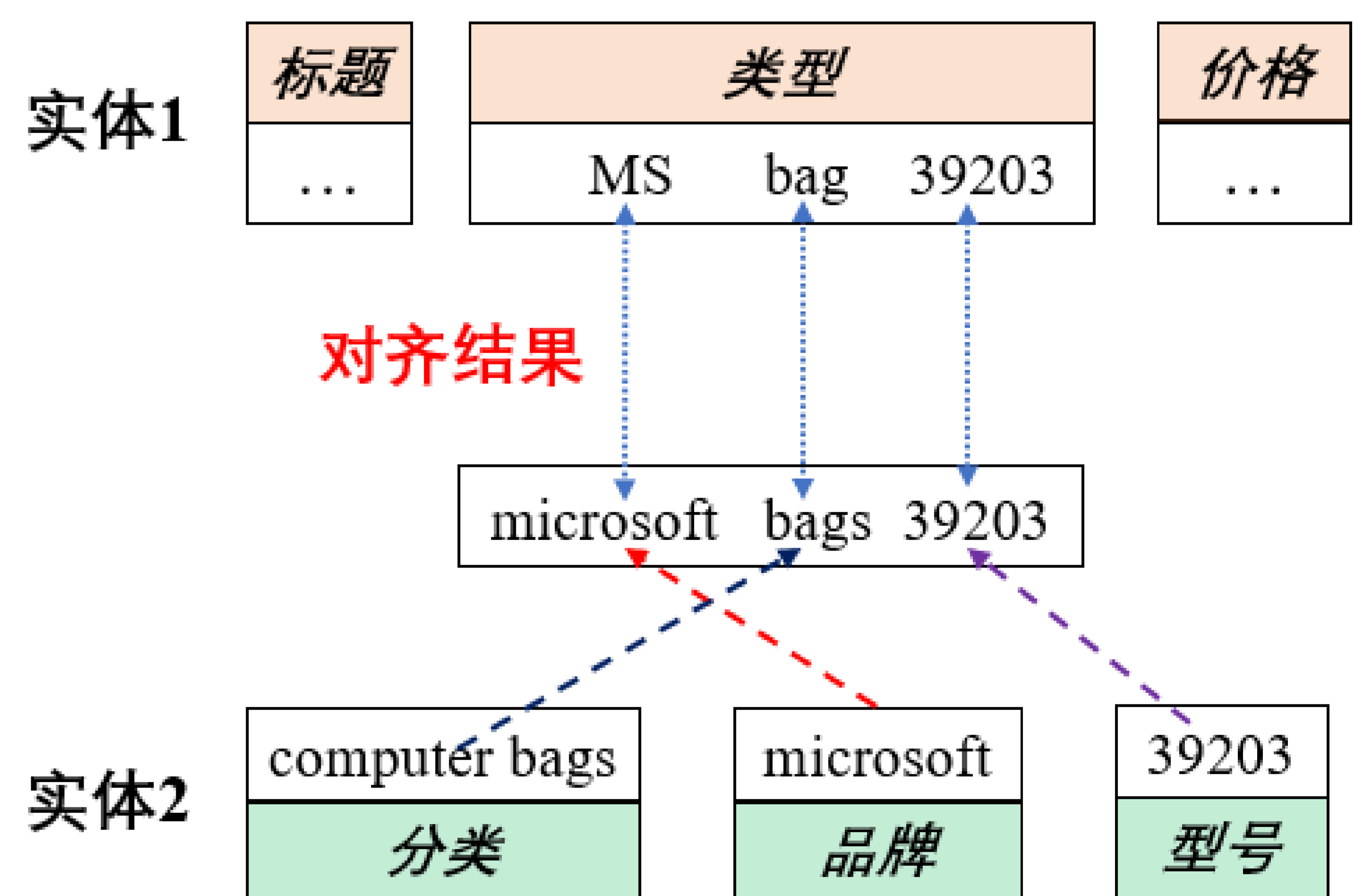
➤ 统一建模多个层次的匹配

➤ 逐层聚合匹配证据 (符号->属性->实体)



跨属性的符号对齐

➤ 跨属性地为每个符号选择匹配对象



属性感知的注意力机制

➤ 自适应地为属性识别重要信息

➤ 聚合过程中, 重要的符号获得更大权重

