ỨNG DỤNG LUẬT KẾT HỢP KHAI PHÁ DỮ LIỆU HỖ TRỢ ĐỊNH HƯỚNG VIỆC LÀM

■ ThS. NGÔ VĂN LINH (*)

TÓM TẮT

Sự phát triển của công nghệ thông tin (CNTT) và ứng dụng CNTT ở hầu hết các lĩnh vực đồng nghĩa với lượng dữ liệu đã được thu thập và lưu trữ ngày càng lớn. Các hệ quản trị cơ sở dữ liệu (CSDL) truyền thống chỉ khai thác được một lượng thông tin nhỏ không còn đáp ứng đầy đủ những yêu cầu, những thách thức mới. Trong phạm vi bài báo này, tác giả nghiên cứu kỹ thuật phát hiện tri thức trong CSDL việc làm và hỗ trợ định hướng xu hướng việc làm cho sinh viên Trường Đại học Kinh tế Công nghiệp Long An.

***Từ khóa**: Khai phá dữ liệu, luật kết hợp, Data Mining, Association Rule, Thuật toán Apriori*

SUMMARY

The development of information technology (IT) and IT applications in almost every field means the amount of data collected and stored is increasing. Traditional database management systems (databases) only exploit a small amount of information that no longer meets new requirements and challenges. In this paper, the author researches the technique of knowledge discovery in the job database and supports the employment trend orientation for students of Long An University of Economics and Industry.

***Key words**: Data mining, association law, Data Mining, Association Rule, Apriori Algorithm*

1. Giới thiệu

Cùng với việc tăng không ngừng khối lượng dữ liệu, các hệ thống thông tin cũng được chuyên môn hóa, phân hoạch theo các lĩnh vực ứng dụng như sản xuất, tài chính, buôn bán thị trường v.v... Như vậy, bên cạnh chức năng khai thác dữ liệu có tính chất tác nghiệp, sự thành công trong kinh doanh không còn là năng suất của các hệ thống thông tin nữa mà là tính linh hoạt và sẵn sàng đáp lại những yêu cầu trong thực tế, CSDL cần đem lại những "tri thức" hơn là chỉnh những dữ liệu đó.

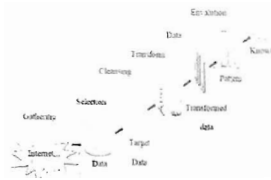
Lúc này các mô hình CSDL truyền thống và ngôn ngữ SQL đã cho thấy không có khả năng thực hiện công việc này. Để lấy được tri thức trong khối dữ liệu khổng lồ này, người ta đã đi tìm những kỹ thuật có khả năng hợp nhất các dữ liệu từ các hệ thống giao dịch khác nhau, chuyển đổi thành một tập hợp các cơ sở dữ liệu ổn định, có chất lượng, chỉ được sử dụng riêng cho một vài mục đích nào đó.

Với những thách thức như vậy, các nhà nghiên cứu đã đưa ra một phương pháp mới trên kho dữ liệu đáp ứng cả nhu cầu trong khoa học cũng như trong hoạt động thực tiễn. Đó chính là công nghệ phát hiện tri thức từ CSDL.

2. Quá trình khai phá dữ liệu

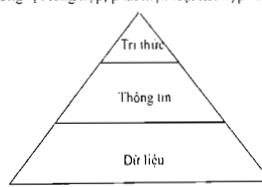
Một vấn đề rất quan trọng để dẫn đến thành công là việc biết sử dụng thông tin một cách có hiệu quả. Điều đó có nghĩa là từ các dữ liệu sẵn có phải tìm ra những thông tin tiềm ẩn có giá trị mà trước đó chưa được phát hiện, phải tìm ra những xu hướng phát triển và những yếu tố tác động lên chúng. Thực hiện công việc đó chính là thực hiện quá trình phát hiện tri thức trong cơ sở dữ liệu (Knowledge Discovery in Database – KDD) mà trong đó kỹ thuật này cho phép ta lấy được các tri thức chính là pha khai phá dữ liệu (KPDL).

(*) Trường ĐH KTCN Long An



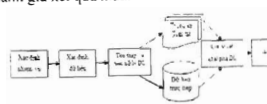
Hình 1: Quá trình khai phá tri thức

Quá trình khám phá tri thức từ CSDL là một quá trình có sử dụng nhiều phương pháp và công cụ tin học nhưng vẫn là một quá trình mà trong đó con người là trung tâm. Quá trình này gồm nhiều giai đoạn [1]. Đầu ra của giai đoạn này là đầu vào của giai đoạn sau. Trong tiến trình này, người ta đặc biệt quan tâm đến pha khai phá dữ liệu (Data Mining) KPDL chính là sử dụng những kỹ thuật, những phương pháp để đưa ra những thông tin có cấu trúc, những tri thức tiềm ẩn trong lượng dữ liệu. Các kỹ thuật phát hiện tri thức được thực hiện qua nhiều giai đoạn và sử dụng nhiều phương pháp như: phân lớp, gom cụm, phân tích sự tương tự, tổng hợp, phát hiện luật kết hợp và mẫu tuần tự,..



Hình 2: Mối quan hệ giữa thông tin, dữ liệu và tri thức

Quá trình phát hiện tri thức gồm các bước cơ bản sau: trích lọc dữ liệu, tiền xử lý dữ liệu, chuyển đổi dữ liệu, KPDL, đánh giá kết quả mẫu.



Hình 3: Quá trình khai phá dữ liệu

- **Trích lọc dữ liệu (Selection):** Đây là giai đoạn tập hợp các dữ liệu được khai thác từ một CSDL, một kho dữ liệu, thậm chí từ các nguồn ứng dụng web vào một CSDL riêng.

- **Tiền xử lý dữ liệu (Preprocessing)**: Phần lớn các CSDL đều ít nhiều mang tính không nhất quán. Vì vậy khi gom dữ liệu rất có thể mắc một số lỗi như dữ liệu không đầy đủ, chặt chẽ và không logic (bị trùng lặp, giá trị bị sai lệch....).

- **Chuyển đổi dữ liệu (Transformation)**: Dữ liệu sẽ được chuyển đổi về dạng thuận tiện để tiến hành các thuật toán khám phá dữ liệu.

- **Khai phá dữ liệu (Data Mining)**: Sử dụng các kỹ thuật nhằm phát hiện ra các tri thức tiềm ẩn trong dữ liệu. Một số kỹ thuật được sử dụng đó là: phân lớp, gom cụm, luật kết hợp,...

- **Đánh giá kết quả mẫu (Evaluation of Result)**: Các mẫu dữ liệu được chiết xuất bởi các phần mềm KPDL. Không phải bất cứ mẫu nào cũng đều có ích, thậm chí còn bị sai lệch. Chính vì vậy, cần phải xác định và lựa chọn những tiêu chuẩn đánh giá sao cho sẽ chiết xuất ra các tri thức cần thiết.

3. Các phương pháp, kỹ thuật khai phá dữ liệu

Các kỹ thuật KPDL có thể được chia làm 2 nhóm chính [2]:

+ Kỹ thuật KPDL mô tả gồm các phương pháp: phân nhóm (Clustering), tổng hợp hóa (Summerization), phát hiện sự biến đổi và độ lệch (Change and deviation detection), phân tích luật kết hợp (Association Rules). ..

+ Kỹ thuật KPDL dự đoán gồm các phương pháp: phân lớp (Classification: Decision Tree, K-Nearest Neighbor, Neural Network, Genetic Algorithms, Bayesian Network, Rough and Fuzzy Sets), hồi quy (Regression), ...

Các loại dữ liệu có thể được khai phá như sau:

+ CSDL quan hệ (Relational Databases): là những CSDL được tổ chức theo mô hình quan hệ. Hiện nay, các hệ quản trị CSDL đều hỗ trợ mô hình này như: MS Access, MS SQL Server, Oracle, IBM DB2,...

+ CSDL đa chiều (Multidimention Structures, Data Warehouse, Data Mart): dữ liệu được chọn từ nhiều nguồn khác nhau và chứa những đặc tính lịch sử thông qua thuộc tính thời gian tường minh hoặc ngầm định.

+ CSDL giao tác (Transaction Databases): là loại dữ liệu được sử dụng nhiều trong siêu thị, thương mại, ngân hàng,...

+ CSDL quan hệ – hướng đối tượng (Object Relational Databases): mô hình CSDL này là lai giữa mô hình hướng đối tượng và mô hình CSDL quan hệ.

+ CSDL không gian và thời gian (Spatial, Temporal, Time – Series Data): chứa những thông tin về không gian địa lý hoặc thông tin theo thời gian.

+ CSDL đa phương tiện (Multimedia Database): là loại dữ liệu có nhiều trên mạng, bao gồm các loại như âm thanh, hình ảnh, video, văn bản và nhiều kiểu dữ liệu định dạng khác.

KPDL có nhiều ứng dụng trong thực tế:

+ Điều trị y học và chăm sóc y tế: một số thông tin về chẩn đoán bệnh lưu trong các hệ thống quản lý bệnh viện. Phân tích mối liên hệ giữa triệu chứng bệnh, chẩn đoán và phương pháp điều trị (chế độ dinh dưỡng, thuốc,...).

+ Sản xuất và chế biến: qui trình, phương pháp chế biến và xử lý sự cố Text mining & Web mining: phân lớp văn bản và các trang web, tóm tắt văn bản,...

+ Lĩnh vực khoa học: quan sát thiên văn, dữ liệu gene, dữ liệu sinh vật học, tìm kiếm, so sánh các hệ gene và thông tin di truyền, mối liên hệ gene và các bệnh di truyền,...

+ Lĩnh vực khác: viễn thông, môi trường, thể thao, âm nhạc, giáo dục,...

4. Luật kết hợp

Luật kết hợp là dạng luật biểu diễn tri thức ở dạng tương đối đơn giản. Mục tiêu của phương pháp này là phát hiện và đưa ra các mối liên hệ giữa các giá trị dữ liệu trong CSDL. Mẫu đầu ra của giải thuật KPDL là tập luật kết hợp tìm được.

Thông tin mà dạng luật này đem lại rất có lợi trong các hệ hỗ trợ ra quyết định. Tìm kiếm được những luật kết hợp đặc trưng và mang nhiều thông tin từ CSDL tác nghiệp là một trong những hướng tiếp cận chính của lĩnh vực khai phá dữ liệu.

Từ khi nó được giới thiệu từ năm 1993, bài toán khai thác luật kết hợp nhận được rất nhiều sự quan tâm của nhiều nhà khoa học. Ngày nay việc khai thác các luật như thế vẫn là một trong những phương pháp khai thác mẫu phổ biến nhất trong việc khám phá tri thức và khai thác dữ liệu (KDD: Knowledge Discovery and Data Mining) [3].

Một cách ngắn gọn, một luật kết hợp là một biểu thức có dạng: X => Y, trong đó X và Y là tập các trường gọi là item. Ý nghĩa của các luật kết hợp khá dễ nhận thấy: Cho trước một cơ sở dữ liệu D là tập các giao tác - trong đó mỗi giao tác T thuộc D là tập các item - khi đó X => Y diễn đạt ý nghĩa rằng bất cứ khi nào giao tác T có chứa X thì chắc chắn T có chứa Y. Độ tin cậy của luật (rule confidence) có thể được hiểu như xác suất điều kiện p(Y thuộc T | X thuộc T). Ý tưởng của việc khai thác các luật kết hợp có nguồn gốc từ việc phân tích dữ liệu mua hàng của khách và nhận ra rằng "Một khách hàng mua mặt hàng x1 và x2 thì sẽ mua mặt hàng y với xác suất là c%".

Các khái niệm cơ bản trong Luật kết hợp:

+ **CSDL giao tác** (Transaction Database): CSDL giao tác D gồm các giao dịch T là tập các giao dịch $t_1, t_2, \ldots, t_n$. $T = \{t_1, t_2, \ldots, t_n\}$. T gọi là cơ sở dữ liệu giao tác (Transaction Database).

Mỗi giao tác t_i bao gồm tập các đối tượng I (gọi là itemset) $I = \{i_1, i_2, \ldots, i_m\}$. Một itemset gồm k items gọi là k-itemset.

+ **Hạng mục (Item)**: thuộc tính nào đó của đối tượng đang xét trong CSDL (i_k: $k \in m$, với m là số thuộc tính của đối tượng).

+ **Tập các hạng mục (Itemset)**: Tập các hạng mục $I = \{i_1, i_2, \ldots, i_m\}$ là tập hợp các thuộc tính của đối tượng đang xét trong CSDL.

+ **Giao tác (Transaction)**: Là tập các hạng mục trong cùng một đơn vị tương tác, mỗi giao tác được xử lý một cách nhất quán mà không phụ thuộc vào các giao tác khác.

+ **Luật kết hợp (Association Rules)**: Là một mối quan hệ điều kiện giữa hai tập các hạng mục dữ liệu X và Y theo dạng sau: Nếu X thì Y, và ký hiệu là $X \rightarrow Y$. Trong đó, X,Y là các tập hạng mục, $X, Y \subseteq I$ và $X \cap Y = \emptyset$. X được gọi là tiền đề và Y được gọi là hệ quả của luật.

- Gọi $I = \{i_1, i_2, \ldots, i_m\}$ là tập các trường gọi là items.
- D là tập giao tác, ở đó mỗi giao tác T, là tập các item $T_i \subseteq I$.
- Ta gọi một giao tác T chứa X nếu $X \subseteq T$ (Với $X \subseteq I$)

Mỗi giao tác T_i có chỉ danh là TID.

+ **Độ hỗ trợ (Support)**: Độ hỗ trợ của một tập hợp X trong cơ sở dữ liệu D là tỷ số giữa các bản ghi TD có chứa tập X và tổng số bản ghi trong D (hay là phần trăm của các bản ghi trong D có chứa tập hợp X), ký hiệu là support(X) hay supp(X).

$$Supp(X) = \frac{|\{T \subset D : Y \supset X\}|}{|D|}$$

Ta có: $0 \leq \text{supp}(X) \leq 1$ với mọi tập hợp X.

Độ hỗ trợ của một luật kết hợp $X \rightarrow Y$ là tỷ lệ giữa số lượng các bản ghi chứa tập hợp XUY, so với tổng số các bản ghi trong D. Ký hiệu: supp($X \rightarrow Y$).

$$Supp(X \rightarrow Y) = \frac{|\{T \subset D : T \supseteq X \cup Y\}|}{|D|}$$

+ **Tập phổ biến (Pattern)**: tập các hạng mục có độ hỗ trợ thoả mãn độ hỗ trợ tối thiểu (minsupp - là một giá trị do người dùng xác định trước). Nếu tập mục X có Supp(X) ≥ Minsupp thì ta nói X là một tập các mục phổ biến.

+ **Độ tin cậy (Confidence)**: Độ tin cậy của một luật kết hợp X → Y là tỷ lệ giữa số các bản ghi trong D chứa X ∪ Y với số bản ghi trong D có chứa tập hợp X. Ký hiệu độ tin cậy của một luật là Conf(R). Ta có 0 ≤ conf(R) ≤ 1.

Nhận xét: Độ hỗ trợ và độ tin cậy có xác suất sau:

Supp(X→Y) = P(X ∪ Y).

Conf(X→ Y) = P(Y / X) = Supp(X ∪ Y) / Supp(X).

Thuật toán Apriori KPDL bằng luật kết hợp:

Thuật toán khai thác các tập phổ biến bằng cách thực hiện nhiều lần duyệt CSDL. Duyệt lần thứ nhất để tính độ phổ biến của các 1-itemset và xác định các item phổ biến từ chúng, nghĩa là độ phổ biến thỏa ngưỡng phổ biến tối thiểu. Trong các lần duyệt sau, thuật toán sẽ kết hợp các itemset phổ biến đã tìm được trong lần duyệt trước để tìm các tập ứng viên. Sau đó, tính độ phổ biến thực sự của các tập ứng viên này nhằm xác định itemset nào trong các tập ứng viên là tập phổ biến thực sự. Các itemset này trở thành các hạt giống cho lần duyệt tiếp theo. Quá trình này thực hiện cho đến khi không còn một tập phổ biến mới nào nữa được sinh ra.

Quy ước: Giả sử các item trong mỗi giao dịch được lưu giữ theo thứ tự từ điển. L_k là tập các k-itemset phổ biến. C_k là tập các ứng viên có k-itemset. Mỗi phần tử của L_k và C_k có 2 thành phần: itemset và độ phổ biến tương ứng

Thuật toán Apriori được mô tả như sau:

```
Input: CSDL giao dịch D và ngưỡng độ phổ biến minSupCount.
Output: F1 chứa danh sách các tập phổ biến trong D thỏa minSupCount.
1) L_1  { j ∈ I   δ (j) ≥minSupCout}
2) for ( k = 2; Lk-1≠ Ø; k++) do
3) Ck – Apriori_gen(Lk-1)
4) for each t ∈ D do
5) for each ck∈ Ck do
6) if ck ⊆ t then ck.count++
7)      Lk   {ck∈ Ck | ck.count ≥minSupCout}
8) F1 = UkLk

Apriori_gen(Lk-1)
9) Ck = Ø
10) for each l1 ∈ Lk-1 do
11) for each l2 ∈ Lk-1 do
12) if (l1[1]   l2[1]) ∧ (l1[2] = l2 [2]) ∧ ... ∧(l1[k-1]   l2[k-1]) then
```

```
13)                    c    l1U l2 // Bước kết hợp l1 và l2 sinh ra ứng
viên c
14) if (Has_infrequent_subset(c, Lk-1)= False then
15)                        Add s into Ck
16) return Ck

Has_infrequent_subset(c, Lk-1)
17) foreach (k -1)-itenset s ⊂ c do
18) if s ∉ Lk-1then
19) return True
20)return False
```

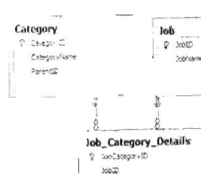
Việc khai thác các luật kết hợp từ CSDL chính là việc tìm tất cả các luật có độ hỗ trợ và độ tin cậy do người sử dụng xác định trước. Các ngưỡng của độ hỗ trợ và độ tin cậy được ký hiệu là *minsup* và *minconf*.

5. Ứng dụng luật kết hợp khai phá dữ liệu việc làm

Sự bùng nổ của công nghệ thông tin, nhu cầu tuyển dụng trực tuyến trở nên phù hợp hơn với các ứng viên và các nhà tuyển dụng so với cách tuyển dụng truyền thống. Với cách tuyển dụng này các ứng viên hay nhà tuyển dụng chỉ cần truy cập vào các Website tuyển dụng tìm các công việc, hay các hồ sơ ứng viên phù hợp với khả năng của các ứng viên hay nhà tuyển dụng và các ứng viên sẽ hộp hồ sơ trực tiếp qua email cho các nhà tuyển dụng. Với cách tuyển dụng mới này cũng giúp cho các nhà quản lý đỡ mất thời gian trong việc thu thập thông tin về việc làm của các cơ quan quản lý, có thể nắm bắt được nhu cầu việc làm của xã hội và có thể từ các thông tin việc làm trong CSDL việc làm có thể rút ra các tri thức hay các xu hướng công việc và là nguồn thông tin giúp Trường Đại học Kinh tế Công nghiệp Long An xác định xu hướng ngành nghề, góp phần định hướng đào tạo của trường.

Hiện nay, do nhu cầu của xã hội, việc tuyển dụng trên các website tuyển dụng khá phổ biến, các thông tin việc tìm người và người tìm việc được cập nhật liên tục. Các thông tin về việc tìm người bao gồm: Ngành tuyển, doanh nghiệp cần tuyển, công việc, mức lương, độ tuổi, giới tính. Các thông tin về người tìm việc bao gồm: Ngành tuyển, người tuyển, độ tuổi, giới tính, công việc. Các thông tin tổng hợp này sẽ giúp các nhà quản lý, các trường đại học biết được xu hướng tuyển của doanh nghiệp, xu hướng chọn ngành nghề của người học, đánh giá về mức lương của mỗi ngành qua đó có điều chỉnh cho phù hợp…

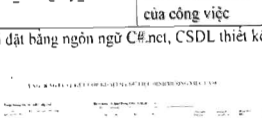
Sơ đồ logic dữ liệu: Công việc được lưu vào bảng Job, mỗi công việc có những thuộc tính: ngành nghề, tỉnh thành, mức lương, trình độ, giới tính, loại hình công việc, kinh nghiệm làm việc…. các thuộc tính này được lưu trong bảng Category.



Hình 4: Sơ đồ Logic CSDL quản lý việc làm

Tên	Ý nghĩa
Category	Danh mục thuộc tính công việc
Job	Danh sách công việc
Job_Category_Details	Chi tiết các thuộc tính của công việc

Chương trình được cài đặt bằng ngôn ngữ C#.net, CSDL thiết kế trên SQL Server 2012, hệ điều hành Window 10.



Hình 5: Giao diện chính của chương trình

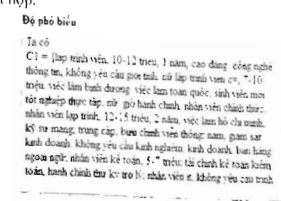
Hình 6: Giao diện import thông tin việc làm

Chuyển dữ liệu sang dạng nhị phân thuận lợi cho việc thao tác tổng hợp dữ liệu việc làm:

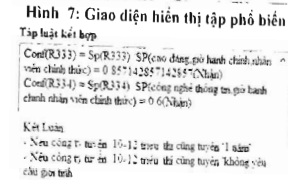
Bước 1: Sau khi kết thúc bước nhập liệu, chọn button "**Biểu diễn dưới dạng nhị phân**" để chuyển CSDL sang dạng ma trận nhị phân.

Bước 2: nhập thông tin minsupp ($0 < minsupp < 100$) và minconf ($0 < minconf <= 100$).

Bước 3: chọn button "**Phân tích**" để tiến hành phân tích bài toán và kết qua thu được chính là tập phổ biến và tập luật kết hợp.



Hình 7: Giao diện hiển thị tập phổ biến



Hình 8: Giao diện hiển thị tập luật kết hợp

Với tập dữ liệu được thu thập từ trang giới thiệu việc làm (http://timviecnhanh.com), đã thu thập được trên 2500 việc làm.

Hình 9: Giao diện trang chủ giới thiệu việc làm

Với thông số độ hỗ trợ Minsupp = 20% và độ tin cậy Minconf = 50%.

Tập phổ biến thỏa điều kiện là:

- **F1** = { {1 năm}; {cao đẳng}; {công nghệ thông tin}; {không yêu cầu giới tính}; {7-10 triệu}; {nữ}; {giờ hành chính}; {nhân viên chính thức}; {việc làm hồ chí minh}; {trung cấp}; {nam}; {tài chính/kế toán/kiểm toán} }
- **F2** = { {1 năm,cao đẳng}; {1 năm,công nghệ thông tin}; {1 năm,không yêu cầu giới tính}; {1 năm,7-10 triệu}; {1 năm,việc làm hồ chí minh}; {cao đẳng,công nghệ thông tin}; {cao đẳng,không yêu cầu giới tính}; {cao đẳng,7-10 triệu}; {công nghệ thông tin,không yêu cầu giới tính}; {không yêu cầu giới tính,việc làm hồ chí minh}; {việc làm hồ chí minh,trung cấp}}

- **F3 = ;;**

Luật kết hợp thỏa điều kiện là:

- R1: 1 năm --> cao đẳng
- R2: cao đẳng --> 1 năm
- R3: 1 năm --> công nghệ thông tin
- R4: công nghệ thông tin --> 1 năm
- R5; 1 năm --> không yêu cầu giới tính
- R6: không yêu cầu giới tính --> 1 năm
- R7: 1 năm --> 7-10 triệu
- R8: 7-10 triệu --> 1 năm
- R9: 1 năm --> việc làm hồ chí minh
- R10: việc làm hồ chí minh --> 1 năm
- R11: cao đẳng --> công nghệ thông tin
- R12: công nghệ thông tin --> cao đẳng
- R13: cao đẳng --> không yêu cầu giới tính
- R14: không yêu cầu giới tính --> cao đẳng
- R15: cao đẳng --> 7-10 triệu
- R16: 7-10 triệu --> cao đẳng
- R17: công nghệ thông tin --> không yêu cầu giới tính
- R18: không yêu cầu giới tính --> công nghệ thông tin
- R19: không yêu cầu giới tính --> việc làm hồ chí minh
- R20: việc làm hồ chí minh --> không yêu cầu giới tính
- R21: việc làm hồ chí minh --> trung cấp
- R22: trung cấp --> việc làm hồ chí minh

Các tri thức thu được phục vụ cho việc định hướng việc làm đối với sinh viên Trường Đại học Kinh tế Công nghiệp Long An:

- Công việc yêu cầu tối thiểu trình độ cao đẳng có 1 năm kinh nghiệm.
- Ngành Công nghệ thông tin yêu cầu có ít nhất 1 năm kinh nghiệm.
- Có ít nhất 1 năm kinh nghiệm thì mức lương từ 7 – 10 triệu.
- Ngành Công nghệ thông tin yêu cầu có trình độ tối thiểu là cao đẳng.
- Ngành Công nghệ thông tin không yêu cầu giới tính.
- Việc làm chủ yếu tập trung khu vực thành phố Hồ Chí Minh.

6. Kết luận

Bài báo nghiên cứu lý thuyết tổng quan về kỹ thuật KPDL dựa trên luật kết hợp. Trên cơ sở đó, tác giả tiến hành thử nghiệm KPDL trên 2500 việc làm được rút trích từ trang http://timviecnhanh.com sử dụng công nghệ C# và SQL Server áp dụng thuật toán Apriori của luật kết hợp KPDL thu thập được, từ đó phân tích nhu cầu tuyển dụng của doanh nghiệp, định hướng việc làm cho sinh viên Trường Đại học Kinh tế Công nghiệp Long An.

Tài liệu tham khảo

[1]. Hoàng Văn Kiếm (2002), *Các hệ cơ sở tri thức*, Giáo trình Công nghệ tri thức và ứng dụng, Đại học Quốc gia TP.HCM.

[2]. Hoàng Văn Kiếm (2014), *Chuyên đề Công nghệ tri thức và ứng dụng*. Bài giảng Công nghệ tri thức và ứng dụng, Trường Đại học Công nghệ thông tin.
[3]. Trần Hùng Cường, Ngô Đức Vĩnh (2011), *Tổng quan về phát hiện tri thức và khai phá dữ liệu*. Tạp chí khoa học, Trường Đại học Công nghiệp Hà Nội.

Ngày nhận:01/8/2018
Ngày duyệt đăng:01/01/2019