

我々の発表では、以下の2件の質問をいただきました。  
発表者の力不足で、いい回答は出来ませんでしたが、どちらも非常に重要な質問(&コメント)だったと思います。  
質問をいただいたお二方と、座長の横田先生(東工大)に感謝いたします。

-----

質問1: 中野美由紀様(東京大学生産技術研究所)  
「発見されたHTMLなどの『構造』は、どのような意味を持つのか?」

回答:  
確かに、HTMLデータを用いた実験では、ブラウザで容易に確認できる繰り返しの構造しか発見されず、マイニングという観点ではあまり面白く感じられないかも知れません。  
今後は、Movie DatabaseなどのXMLデータを用いて実験をおこない、面白いパターンを発見できればと思います。

備考:  
この説明では、中野様にはご納得いただけませんでした。  
Movie Databaseが関係データベースだと思われてしまったかも知れません。  
あるいは、半構造データから(原子値をもたない)構造を発見することの意義を質問されたのかも知れません。  
どちらにせよ、この質問は我々の議論が足りない部分であると感じています。

-----

質問2: 鳥居修様(東芝)  
「候補パターンを生成する際に、すべてのラベルに関して生成するのは効率が悪いのでは?」

回答:  
我々の基本アルゴリズムでは、すべてのラベルに関して候補パターンを作成していますが、実際には高速化の部分で説明した手法(非頻出1-パターンと非頻出2-パターンを用いた高速化)を導入しています。

備考:  
こちらにも、鳥居様には納得していただけませんでした。後から気づいたのですが、我々の手法より効率的な候補パターン生成法がZakiにより提案されています。  
そのことを指摘したうえで、Zakiの手法は我々のアルゴリズムにも適用可能であるが、適用は今のところ行っていないことをご説明申し上げれば、納得していただけたかも知れません。