

Market Basket Analysis untuk Swalayan KSU Sumber Makmur dengan Algoritma FP Growth

Ramadhan, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya*,
Esther Irawati Setiawan, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya*.

Abstrak—Salah satu teknik data mining yang populer digunakan adalah association data mining atau yang biasa disebut dengan istilah market basket analysis. Market basket didefinisikan sebagai suatu itemset yang dibeli secara bersamaan oleh pelanggan dalam suatu transaksi. Market basket analysis adalah suatu sarana untuk meningkatkan penjualan. Metode ini dimulai dengan mencari sejumlah frequent itemset dan dilanjutkan dengan pembentukan aturan-aturan asosiasi. Algoritma Apriori dan frequent pattern growth adalah dua algoritma yang sangat populer untuk menemukan sejumlah frequent itemset dari data-data transaksi yang tersimpan dalam basis data. Dalam penelitian ini algoritma frequent pattern growth (FP Growth) digunakan untuk menemukan sejumlah aturan asosiasi dari basis data transaksi penjualan di Swalayan KSU Sumber Makmur (Trenggalek). Dari hasil pengolahan data didapatkan pola pembelian paling kuat berupa jika membeli pasta gigi maka dimungkinkan juga akan membeli sabun dan jika membeli shampo juga akan membeli sabun dengan tingkat keyakinan (confidence) 63% dan 62%.

Kata Kunci—data mining, association rules, market basket analysis, Apriori, FP-growth.

I. PENDAHULUAN

Sistem komputerisasi dapat membantu pencatatan dan pengolahan data transaksi pembelian dan penjualan barang di toko swalayan atau supermarket. Penggunaan sistem komputerisasi tersebut umumnya terbatas untuk pengelolaan barang. Swalayan KSU Sumber Makmur – Trenggalek telah menggunakan sistem komputerisasi untuk mencatat transaksi penjualan tetapi belum dimanfaatkan untuk mendapatkan *knowledge* untuk meningkatkan transaksi penjualan.

Market Basket Analysis merupakan sebuah analisis terhadap kebiasaan customer berbelanja pada supermarket dengan cara menemukan asosiasi dan korelasi di antara berbagai macam item yang dimasukkan customer di dalam keranjang belanjanya. Market Basket Analysis paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, desain katalog, dan proses pembuatan keputusan bisnis.

Ramadhan, Departemen Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: ramokdr@gmail.com)

Esther Irawati Setiawan, Departemen Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: esther@stts.edu)

Produk-produk yang sering dibeli secara bersamaan dapat ditempatkan secara berdekatan sehingga konsumen dapat dengan mudah menemukan apa yang dicari. Dengan demikian para konsumen akan merasa puas dan penjualan juga akan meningkat. *Market Basket Analysis* bertujuan untuk mengetahui *item* apa saja yang sering dibeli bersamaan oleh customer. *Item* di sini diartikan sebagai berbagai macam produk atau barang pada supermarket.

Banyak algoritma yang bisa digunakan untuk melakukan *mining frequent itemset* untuk mengetahui hubungan antar *item* dalam transaksi penjualan. Apriori adalah algoritma pertama dan menjadi dasar dalam pencarian *frequent itemset* yang diusulkan oleh R. Agrawal dan R. Srikant di tahun 1994 [1]. FP Growth merupakan algoritma yang saat ini dikenal cepat dan umum digunakan tanpa pembangkitan *candidate generation*. FP Growth meringkas informasi transaksi ke dalam bentuk pohon.

II. TINJAUAN PUSTAKA

A. Market Basket Analysis

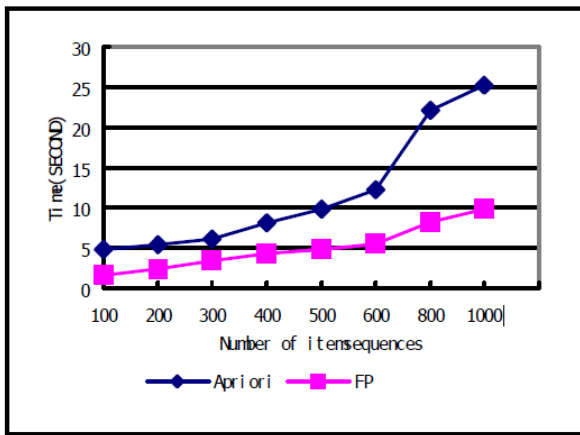
Pada paper [3] disebutkan bahwa *market basket analysis* atau analisis keranjang belanja memiliki tujuan untuk mengidentifikasi produk, atau kelompok produk, yang cenderung terjadi bersamaan (terkait) dalam transaksi pembelian (keranjang). Pengetahuan yang diperoleh dari analisis keranjang belanja bisa sangat berharga; misalnya, dapat digunakan oleh supermarket untuk mengatur ulang tata letaknya, mengambil produk yang sering dijual bersama dan menempatkannya berdekatan. Dapat juga digunakan untuk meningkatkan efisiensi promosi: produk yang terkait seharusnya tidak dipromosikan pada saat yang sama. Dengan mempromosikan secara tepat salah satu produk terkait memungkinkan meningkatkan penjualan produk itu dan mendapatkan penjualan yang menyertainya peningkatan untuk produk terkait. Basis data Untuk analisis keranjang belanja terdiri dari semua transaksi yang dilakukan dalam periode penjualan tertentu (misalnya satu tahun) dan di lokasi penjualan tertentu (misalnya di suatu lokasi supermarket). Konsumen bisa tampil lebih dari sekali di database. Bahkan, konsumen akan muncul didatabase setiap kali mereka melakukan transaksi di lokasi penjualan. Tujuan dari analisis ini adalah untuk menemukan yang terbaik kombinasi produk yang sering dibeli oleh pelanggan.

Penulis melakukan penelitian pada implementasi metode data mining untuk Six Sigma metodologi. Kemudian dipilih

market basket analysis untuk meningkatkan fase, dengan itu untuk memprediksi kebiasaan pelanggan.

B. Analisis Asosiasi

Pada paper [2] penulis melakukan penelitian tentang seberapa efektif metode FP Growth. Ekperimen dilakukan dengan 2 jenis data yang memiliki 10 ribu dan 100 ribu transaksi. Eksperimen dilakukan menggunakan algoritma Apriori dan FP Growth. Perbandingan waktu proses ditunjukkan oleh Gambar. 1.



Gambar. 1. Perbandingan waktu proses Apriori dan FP Growth

Analisis asosiasi atau *association rules mining* adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi *item*. *Interestingness measure* yang dapat digunakan dalam data mining adalah :

- a. Support, adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi. Nilai support sebuah *item* diperoleh dengan menggunakan rumus berikut:

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ Transaksi}$$

Nilai support dari 2 *item* diperoleh dengan menggunakan rumus :

$$Support(A, B) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi}$$

- b. Confidence, adalah suatu ukuran yang menunjukkan hubungan antar dua *item* secara conditional (berdasarkan suatu kondisi tertentu). Confidence dihitung dengan rumus :

$$P(B|A) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi\ Mengandung\ A}$$

C. Algoritma Apriori

Pada paper [1] penulis memperkenalkan algoritma yang diberi nama Apriori. Langkah pertama dari algoritma Apriori menghitung kemunculan *item* untuk menentukan suatu *itemsets* besar. Langkah berikutnya adalah mencari *k*, yang terdiri dari dua fase. Pertama, *itemsets* besar *L_{k-1}* ditemukan di jalur (*k-1*) digunakan untuk menghasilkan kandidat *itemsets* *ck*, menggunakan fungsi *apriorigen*. Selanjutnya, database dipindai dan dukungan kandidat di *ck* dihitung. Agar lebih cepat perlu langkah efisien untuk menentukan kandidat di *ck* yang terdapat di transaksi yang diberikan *t*.

Bentuk algoritma Apriori dapat dituliskan seperti nampak di Gambar. 2.

```

L1 = {frequent itemset with one element}
for (k=2; Lk-1 ≠ ∅; k++)
{
  Ck = apriori-gen(Lk-1); //pembuatan
                             kandidat
                             //baru
  for all transactions t
  {
    C't = subset(Ck, t); //kandidat yang
                          //tampil pada t
    for all candidates c ∈ C't, do
      c.count ++;
  }
  Lk = {c ∈ C't | c.count ≥ minsup}
}
return ∪k Lk;

```

Gambar. 2. Algoritma Apriori

dimana :

- L : himpunan frequent itemset
- minsup : minimum support
- C : himpunan kandidat itemset
- c : kandidat itemset
- t : transaksi

D. Algoritma FP Growth

Pada paper [4] penulis melakukan perbandingan kinerja algoritma Apriori dengan FP Growth. Algoritma FP Growth membantu menyelesaikan 2 masalah utama pada algoritma apriori, yaitu : menghasilkan sejumlah besar kumpulan kandidat dan berulang kali memindai database. FP Growth diusulkan untuk untuk memberi solusi atas permasalahan di algoritma Apriori dengan menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada algoritma Apriori, yaitu : strategi *divide-and-conquer*. Kelebihan dari algoritma FP Growth adalah hanya melakukan 2 kali pemindaian ke database dan tidak membentuk kumpulan kandidat yang besar. Pertama kali algoritma FP Growth membentuk freuent

pattern tree atau disebut juga FP-Tree. Proses penyusunan FP-Tree dari mulai representasi awal transaksi untuk menentukan frequent itemset, setelah didapatkan frequent itemset kemudian dilanjutkan ke perhitungan *association rule*, menghitung *rule* yang dihasilkan dari penerapan metode FP Growth untuk mendapatkan pola terbaik dari pola-pola yang ada. Hasil akhir pola dari perhitungan *association rule* merupakan hasil rekomendasi sebagai bahan acuan terbaik dalam menentukan item atau produk yang paling sering dibeli oleh konsumen.

Algoritma FP Growth ditunjukkan oleh Gambar. 3.

Algorithm FP-growth

Input: $\mathcal{D}, \sigma, I \subseteq \mathcal{I}$

Output: $\mathcal{F}[I](\mathcal{D}, \sigma)$

```

 $\mathcal{F}[I] := \{\}$ 
for all  $i \in \mathcal{I}$  occurring in  $\mathcal{D}$  do
   $\mathcal{F}[I] := \mathcal{F}[I] \cup \{I \cup \{i\}\}$ 
   $H := \{\}$ ;  $\mathcal{D}^i := \{\}$ 
  for all  $j \in \mathcal{I}$  occurring in  $\mathcal{D}$  such
  that  $j > i$  do
    if  $supp_{\mathcal{D}}(I \cup \{i, j\}) \geq \sigma$  then
       $H := H \cup \{j\}$ 
  for all  $(tid, X) \in \mathcal{D}$  with  $i \in X$  do
     $\mathcal{D}^i := \mathcal{D}^i \cup \{(tid, X \cap H)\}$ 
  Compute  $\mathcal{F}[I \cup \{i\}](\mathcal{D}^i, \sigma)$ 
   $\mathcal{F}[I] := \mathcal{F}[I] \cup \mathcal{F}[I \cup \{i\}]$ 

```

Gambar. 3. Algoritma FP Growth

Metodologi algoritma FP-Growth bisa diperinci sebagai berikut :

1. Menentukan nilai *minimum support*, misalnya ditentukan 40%. Dengan asumsi jumlah transaksi 9 maka nilai *minimum support count* adalah $40/100 \cdot 9 = 3,6$ (dibulatkan ke atas jadi 4).
2. Berikutnya adalah menghitung jumlah kemunculan masing-masing item (*support count*).
3. Selanjutnya cari *item* yang memenuhi *minimum support count*, urutkan *support count* secara *descending* (besar ke kecil) dan *itemset* secara *ascending* (kecil ke besar).
4. Daftar *itemset* dan *support count* itu bisa disebut juga sebagai *header table*, selanjutnya dibuat database transaksi baru yang mana dilakukan pengurutan *list of item* sesuai urutan *support count* sekaligus membuang *item* yang tidak ada di dalam *header table* tersebut.
5. Dengan menggunakan database transaksi baru dilakukan pembuatan FP-Tree, dimulai dari transaksi ke-1, 2, dan seterusnya.
6. Dengan menggunakan FP-Tree yang telah dibangun di tahap sebelumnya, dilakukan pembangkitan *conditional pattern base*.
7. Hitung *support count* dari setiap *item* pada setiap *conditional pattern base*, untuk masing-masing *item* yang memiliki jumlah *support count* lebih besar atau sama dengan *minimum support count* akan dibangkitkan menjadi *conditional FP-Tree*.
8. Berikutnya adalah pencarian *Frequent Itemset*,

apabila *conditional FP-Tree* merupakan lintasan tunggal (*single path*), maka didapatkan *frequent itemset* dengan melakukan kombinasi *item* untuk setiap *conditional FP-Tree*. Jika bukan lintasan tunggal, maka dilakukan pembangkitan FP-Growth secara *rekursif* (proses memanggil dirinya sendiri).

9. Tahap selanjutnya adalah pencarian *strong association rule*. *Strong association rule* adalah *rule* yang memenuhi syarat *minimum support* dan *minimum confidence*. Tahap pencarian *strong association rule* dilakukan dengan langkah :

- a. Untuk setiap *frequent itemset* l dari , dibangkitkan semua *non-empty subset* dari l .
- b. Untuk setiap *non-empty subset* s dari l , menghasilkan *rule* :

$$s \Rightarrow (l-s) \text{ jika } \frac{\text{support_count}(l)}{\text{support_count}(s)} \geq \text{min_conf}$$

dimana *min_conf* adalah *minimum confidence*.

Untuk analisis digunakan dataset *french retail dan bakery shop*. Minimum support ditentukan sebesar 1% dan minimum confidence adalah 50%. Komputer yang digunakan untuk ujicoba memiliki spesifikasi : processor Inter® Core™ i5-4210U, memori utama 4GB, dan operating system Microsoft Windows 10.

Kesimpulan dari penelitian pada paper tersebut menyatakan bahwa FP Growth memerlukan waktu lebih cepat pada segala pengujian, baik saat produk dikurangi maupun tidak.

E. Hybrid-Dimension Association Rules

Pada paper [5] penulis menerapkan konsep multi-dimensional yang mana digunakan 2 faktor atau variable untuk keputusan ini, yaitu : barang yang dibeli bersamaan dan waktu transaksi. Terdapat beberapa langkah untuk melakukan data mining, dimulai dari pre-prosesing, menentukan *frequent itemsets*, menentukan *association rules*.

Percobaan dilakukan menggunakan data transaksi dari minimarket X yang ingin melakukan analisis kebiasaan berbelanja pelanggan.

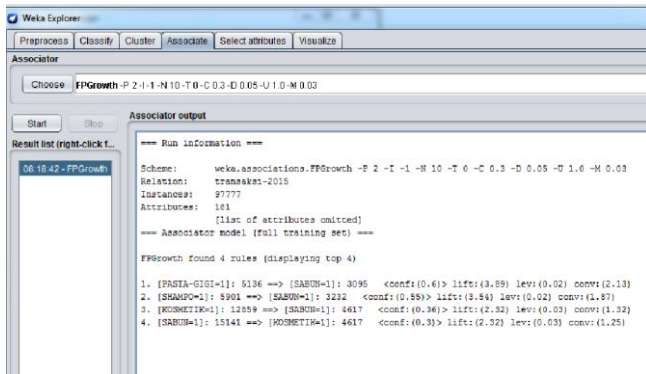
Hasil dari percobaan menunjukkan suatu *rules* bahwa pelanggan berbelanja di hari Minggu untuk membeli *Aqua Isi Ulang* dan *Club Isi Ulang*, yang mana hal ini memenuhi kriteria *Hybrid-Dimension Association Rules*.

F. Perbedaan Algoritma Apriori dan FP-Growth

Secara umum bisa digambarkan bahwa ada beberapa hal yang membedakan antara algoritma Apriori dan algoritma FP-Growth, berikut ini adalah ringkasannya :

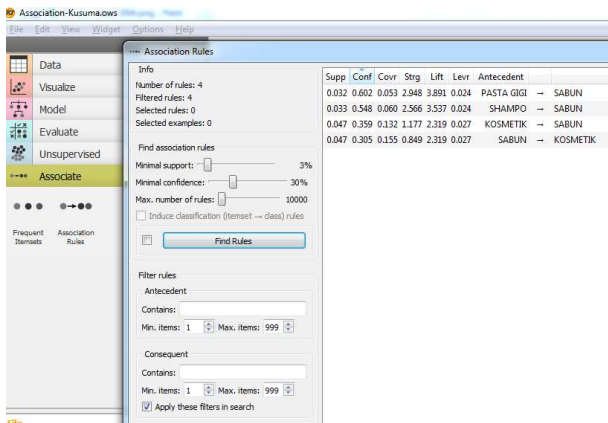
- Kekurangan algoritma Apriori :
- Membutuhkan ruang penyimpanan besar
 - Membutuhkan waktu proses lebih lama
 - Menggunakan memori lebih banyak
 - Kurang akurat
 - Melakukan pembacaan ke database berulang kali

confidence dengan menggunakan data tahun 2015-2019 yang memiliki 218 *attributs* dan 612.439 *instances* ditunjukkan oleh Tabel 1.



Gambar. 5. Contoh *Associator output* aplikasi Weka

Pada aplikasi orange bisa dipilih nilai minimum support dan minimum confidence yang diinginkan di bagian *slider* minimal support dan minimal confidence. Dari pengolahan data dengan aplikasi Orange didapatkan tampilan seperti pada Gambar. 6.



Gambar. 6. Contoh tampilan *Association rules* aplikasi Orange

Agar mendapatkan nilai minimum support yang tepat maka dilakukan percobaan dengan nilai tertentu misalnya 30%. Dengan nilai minimum support 30% ternyata belum menghasilkan *rules* karena variasi barang yang diolah cukup banyak yakni mencapai 218 jenis barang untuk seluruh data antara tahun 2015-2019 dan jumlah *instance/* transaksi adalah 612.349. Dengan dataset semacam itu maka nilai minimum support yang bisa menghasilkan *rules* dan tidak terlalu banyak *rules* yang dihasilkan adalah 0.03 atau 3%.

Nilai minimum confidence juga dilakukan percobaan dengan menggunakan nilai yang cukup rendah dulu, misalnya : 50% atau 0.5 di aplikasi Weka. Dengan kombinasi nilai minimum support 30% dan nilai minimum confidence 50% ternyata hanya didapatkan 2 *rules* sehingga dicoba lagi dengan menurunkan nilai minimum confidence ke 20%, hasilnya didapatkan 12 *rules*. Karena *rules* yang muncul sudah terlalu banyak dan yang terendah bernilai confidence 0.21 atau 21% maka ditentukan menggunakan nilai confidence 30% yang menghasilkan 4 *rules*.

Hasil uji coba berbagai macam nilai *minimum support* dan

TABEL 1
HASIL UJI COBA BERBAGAI MACAM NILAI *MINIMUM SUPPORT* DAN *CONFIDENCE*

Minimum Support (%)	MINIMUM CONFIDENCE (%)	JUMLAH ASSOCIATI ON RULE	NILAI CONFIDENCE TERTINGGI
10	90	0	0
5	90	0	0
1	80	0	0
1	70	2	80
1	60	7	80
1	50	8	80
1	40	20	80
1	30	43	80
1	20	102	80
1	10	166	80
2	10	42	62
3	10	12	63
4	10	4	62
5	10	2	37
6	10	0	0
5	20	2	37
4	30	3	62
3	30	4	63

Perbandingan hasil association rule dari kedua aplikasi tersebut nampak seperti pada Tabel 2. Dari Tabel 2. bisa diketahui jika *rule* terbaik atau menarik adalah :

$$\text{beli PASTA GIGI} \Rightarrow \text{beli SABUN} [\text{support} = 3,7\%, \text{confidence} = 63\%].$$

Itu berarti pelanggan yang membeli pasta gigi dan sabun secara bersamaan terjadi pada 3,7% dari keseluruhan transaksi dan tingkat keyakinan untuk dibeli bersama-sama sebesar 63%. Kenapa dikatakan menarik karena nilai support dan confidence cukup tinggi.

V. KESIMPULAN

Nilai minimum support yang tepat pada penelitian dengan menggunakan data transaksi penjualan di Swalayan KSU Sumber Makmur adalah 0.03 atau 3%. Sedangkan nilai minimum confidence di angka 0.3 atau 30%.

Dari hasil pengolahan data didapatkan 2 *strong association* dengan nilai confidence paling tinggi, yaitu: jika membeli pasta gigi maka ada kemungkinan sebesar 63% akan membeli sabun dan jika membeli shampo maka ada kemungkinan 62% akan membeli sabun.

Pencarian *strong association* menggunakan aplikasi Weka versi 3.8.4 dan Orange versi 3.24.1 menghasilkan *output* yang sama meskipun ada perbedaan cara menampilkan nilai *confidence* yang mana Orange dengan 1 desimal sedangkan Weka tanpa desimal.

TABEL 2

PERBANDINGAN ASSOCIATION RULES HASIL DARI WEKA DAN ORANGE

TAHUN	ANTECEDENT	CONSEQUENT	WEKA		ORANGE
			SUPPORT	CONFIDENCE	CONFIDENCE
2015	PASTA GIGI	SABUN	3,2	60,2	60
	SHAMPO	SABUN	3,3	54,8	55
	KOSMETIK	SABUN	4,7	35,9	36
	SABUN	KOSMETIK	4,7	30,5	30
2016	PASTA GIGI	SABUN	3,1	63,1	63
	SHAMPO	SABUN	3,3	55,9	56
	KOSMETIK	SABUN	4,6	36,5	37
2017	PASTA GIGI	SABUN	4,0	64,0	64
	SHAMPO	SABUN	4,4	62,5	62
	BABY CARE	SABUN	3,2	40,7	41
	KOSMETIK	SABUN	5,1	38,4	38
2018	SHAMPO	SABUN	4,8	66,1	66
	PASTA GIGI	SABUN	4,0	64,5	65
	KOSMETIK	SABUN	5,3	38,5	38
	SABUN	KOSMETIK	5,3	30,1	30
2019	SHAMPO	SABUN	4,7	64,7	65
	PASTA GIGI	SABUN	3,8	62,4	62
	KOSMETIK	SABUN	5,5	36,7	37
	SABUN	KOSMETIK	5,5	32,3	32
	SUSU	SNACK	4,2	30,4	30
2015-2019	PASTA GIGI	SABUN	3,7	63,0	63
	SHAMPO	SABUN	4,2	61,9	62
	KOSMETIK	SABUN	5,1	37,3	37
	SABUN	KOSMETIK	5,1	30,2	30

DAFTAR PUSTAKA

- [1] R. Agrawal, R. Srikant, and others, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, 1994, vol. 1215, pp. 487-499.
- [2] W. Zhang, H. Liao, and N. Zhao, "Research on the FP growth algorithm about association rule mining," in *2008 International Seminar on Business and Information Management*, 2008, vol. 1, pp. 315-318.
- [3] A. Trnka, "Market basket analysis with data mining methods," in *2010 International Conference on Networking and Information Technology*, 2010, pp. 446-450.
- [4] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1-6.
- [5] D. H. Setiabudi, G. S. Budhi, I. W. J. Purnama, and A. Noertjahyana, "Data mining market basket analysis' using hybrid-dimension association rules, case study in Minimarket X," in *2011 International Conference on Uncertainty Reasoning and Knowledge Engineering*, 2011, vol. 1, pp. 196-199.
- [6] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *Morgan Kaufmann Ser. Data Manag. Syst.*, vol. 5, no. 4, pp. 83-124, 2011.