A NEW METHOD FOR ANALYZING CORRELATIONS AMONG CURRENCY EXCHANGE RATES

HAI-ZHOU LI[1], HONG YAN[1,2]

[1]Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong
[2]School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia
E-MAIL: 50096081@student.cityu.edu.hk, h.yan@cityu.edu.hk

Abstract:

The correlation between different currency exchange rates has been studied for many years and a number of techniques have been developed. In this paper, we present a new algorithm to analyze the correlation between exchange rates based on biclustering. This algorithm is comprised of two parts. In the first part, the Fast Hough transform is used to detect the lines in the exchange rate pair space. This phase is also called sub-biclustering and every line identified represents a sub-bicluster. In the second part, the sub-biclusters are combined based on comparison and merging. Experiment results show that this biclustering algorithm is very effective. The bicluster patterns are consistent with the underlying economic reasons.

Keywords:

Currency exchange rates; Time series analysis; Financial data analysis; Biclustering; The Fast Hough transform

1. Introduction

Analysis of correlation among different currency exchange rates is of interest to many people in business [1-3]. This kind of correlation is beneficial for much economic research work, such as accurate financial forecasting. The US dollar is the most important currency today in the international monetary system and it is regarded as a reference for all other currencies in the world. Therefore, the exchange rate mentioned in this paper is the one between a currency and the US dollar. There are several types of relationships between exchange rates and each type has unique characteristics. For example, a desire to maintain the competitiveness of Japanese exports to the United States with German exports to the United States leads the Bank of Japan to intervene to ensure a matching depreciation of the yen against the dollar whenever the Deutsche mark (DM) depreciates against the US dollar [4]. On the other hand, a preference for price stability may lead the Bank of Japan to intervene to ensure a matching appreciation of the yen against the dollar, whenever the DM appreciates against the US dollar. The following is another example. Because of the fixed exchange rate regime strictly implemented in Argentina and Egypt in the 1990s, the Argentine peso versus US dollar and the Egyptian pound versus US dollar exchange rates experience almost no change. The correlation between different currency exchange rates consists of three key factors. The first one is the type of correlation. The second one relates to which exchange rates are involved in the correlation. The third is at which time points these involved exchange rates have this type of correlation. This is a typical biclustering problem. The exchange rates of different currencies can be organized into a two dimensional matrix. The columns of this matrix correspond to the selected exchange rates, and the rows correspond to the selected time points. The exchange rate at a time point can be the real exchange rate value or the difference value. The biclustering algorithm can group a certain set of exchange rates together based on the fact that they have a certain type of defined correlation in a certain set of time points [5].

The term biclustering was first used by Cheng and Church [6] in gene expression data analysis. Biclustering refers to a class of clustering algorithms that perform row-column clustering simultaneously. The detected bicluster pattern contains a subset of whole rows and a subset of whole columns. The concept of biclustering has already been widely applied in other research fields besides bioinformatics since it was first proposed. For example, in document clustering [7], biclustering algorithms are applied to cluster documents and words simultaneously. In this paper, the biclustering method is applied to currency exchange rate data analysis.

A different viewpoint of biclustering is proposed in [8]. It can be called geometric biclustering. It is performed in terms of the spatial geometrical distribution of points in a multi-dimensional data space. Based on the concept of geometric biclustering, we develop a new method for currency exchange rate analysis. In our algorithm, we

1-4244-0973-X/07/$25.00 ©2007 IEEE

choose the Fast Hough transform (FHT) [9] to identify the line patterns in two dimensional spaces, each containing a pair of exchange rates. Each line represents a sub-bicluster. An expansion algorithm is used to combine these sub-biclusters to form the complete biclusters based on comparison and merging.

The paper is organized as follows. Section 2 demonstrates that different biclusters can be formulated using the linear relation in a column-pair space. Section 3 describes our biclustering algorithm based on the FHT algorithm and the expansion algorithm. In Section 4, we present results of several experiments to verify the performance of our biclustering algorithm for the extraction of correlations between different exchange rates. Finally, we conclude the paper in Section 5.

2. Linear pattern representation of biclusters

In this paper, we address five major classes of biclusters:

(1) Biclusters with constant values.
(2) Biclusters with constant rows.
(3) Biclusters with constant columns.
(4) Biclusters with additive coherent values.
(5) Biclusters with multiplicative coherent values.

Actually, constant biclusters are a special case of all other types of biclusters. Constant row and constant column biclusters are special cases of additive and multiplicative biclusters. Thus, we reduce the original five classes of biclusters to two classes, additive and multiplicative ones and only consider these two structures for currency exchange rate analysis in this paper.

The two classes of biclusters both have an important property that any two columns (a column-pair) in the bicluster form a sub-bicluster of the same type. Let us represent a column variable as x_i for exchange rate i. Then additive and multiplicative biclusters can be represented using the linear equation, $x_j = kx_i + c$, where k and c are constants. The equation represents an additive bicluster when k equals to1 and a multiplicative one when c equals 0. Therefore, we can find these linear relations in exchange rate pair spaces first and then combine similar sub-biclusters to identify a complete bicluster, instead of searching through all exchange rates at the same time using the strategy of the original geometric biclustering algorithm [8].

3. Geometric biclustering algorithm

We present a biclustering algorithm based on the concept of geometric biclustering. A similar method has already been used for microarray gene expression data analysis [10]. In the first phase of our biclustering algorithm, the lines are detected in the 2D space corresponding to each exchange rate pair. In the second phase, we use our expansion algorithm to form the complete biclusters. The FHT algorithm is chosen because of its robustness for line detection. The expansion algorithm consists of comparison and mergence. In this section, we first introduce the FHT and then propose the expansion algorithm.

3.1. The Fast Hough transform for line detection

The FHT algorithm is a powerful technique for multi-dimensional pattern detection and parameter extraction. It has advantages when little is known about the pattern [9].

We use x-y to represent the data space, and p-q to represent the parameter space. The FHT algorithm will transform a point in the x-y space into a line in the p-q parameter space. If n points on a line in the x-y data space are known, the line obtained from each such point should intersect at the same point in the p-q space. That intersection point defines the parameters of the line in the x-y space. Figure 1 shows the relation between line and point in their respective space.

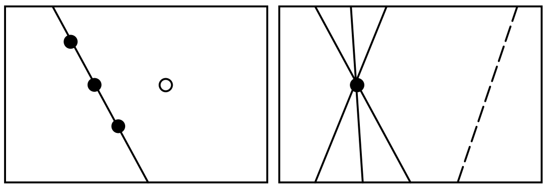

Figure 1. Several points on a line to be detected and a random point in the original data space (left) and lines in the parameter space (right), corresponding to the points in the data space. We can detect the line in the data space by locating the intersection point of the lines in the parameter space.

In the FHT algorithm, points of concentration are detected by dividing the p-q parameter space into a list of accumulators, which are organized into a hierarchical data structure. The FHT algorithm recursively divides the p-q parameter space into squares from low to high resolutions and performs the subdivision and subsequent “vote counting” only on squares with votes exceeding a selected threshold. Each square represents an accumulator. When the division reaches a selected resolution, the left accumulators contain points of concentration. The center coordinates of the square are used as the slope and the intercept of a detected line [9].

The FHT is applied on the 2D space corresponding to an exchange rate pair. A detected line can be described in the form

$$y = kx + c \tag{1}$$

where k is the slope, and c takes the value of the point of intersection with the ordinate axis. Parameters k and c are the key factors that determine the correlation type between the exchange rates. The line detection process can be regarded as sub-biclustering. Every detected line can be regarded as a sub-bicluster. Because the detected sub-bicluster in this step contains two exchange rates, it is called a 2ER sub-bicluster (2 exchange rates sub-bicluster). Each 2ER sub-bicluster comprises all the approximate points on the detected line.

3.2. The expansion algorithm

If the number of selected different exchange rates is N and the number of selected time points is M, the generated exchange rate data is a 2-dimension matrix with size M by N. There are C_N^2 different pair-wise exchange rate combinations in total. In the first phase of our biclustering algorithm, the FHT algorithm is needed for each of the C_N^2 exchange rate pairs. After that, all the potential 2ER sub-biclusters are detected.

Based on the discussion in Section 2, we consider only two types of biclusters. They are biclusters with additive coherent values and biclusters with multiplicative coherent values. The rules for determining the type of a 2ER sub-bicluster are:

(1) If the slope parameter k approximately equals 1, the 2ER sub-bicluster has additive coherent values.

(2) If the slope parameter k does not approximately equal 0 or 1, and at the same time the intersection parameter c approximately equals 0, then the 2ER sub-bicluster has multiplicative coherent values.

Obviously, some of the 2ER sub-biclusters generated from the C_N^2 exchange rate pairs does not fit into either one of the two bicluster types. Therefore, the first step in our expansion algorithm is removing that kind of 2ER sub-biclusters.

Mergence is the main technique used in the expansion algorithm. The regulations for mergence are as follows:

(1) The conditions of merging two 2ER sub-biclusters together are:
   a) The two 2ER sub-biclusters have a common exchange rate.
   b) The two 2ER sub-biclusters have the same type.
   c) The two 2ER sub-biclusters have a reasonable degree of overlapping among the time points they contain.

   Only when all these conditions are satisfied at the same time, can the two 2ER sub-biclusters be merged into a new 3ER sub-bicluster (3 exchange rates sub-bicluster). This new 3ER sub-bicluster comprises the three exchange rates contained in the two 2ER sub-biclusters. The time point set contained in this new 3ER sub-bicluster is the intersection of the time point sets of the two 2ER sub-biclusters. The type of this new 3ER sub-bicluster is the same as the type of those two 2ER sub-biclusters.

(2) Every new nER sub-bicluster (n exchange rates sub-bicluster) is generated by combining a 2ER sub-bicluster and an (n-1)ER sub-bicluster. The conditions for merging a 2ER sub-bicluster and an (n-1)ER sub-bicluster are similar to those of merging two 2ER sub-biclusters:
   a) The 2ER sub-bicluster and the (n-1)ER sub-bicluster have a common exchange rate.
   b) They are of the same type.
   c) They have a reasonable degree of overlapping between the time points they contain.

   In the same way, only when all these three conditions are satisfied at the same time, can the 2ER sub-bicluster and the (n-1)ER sub-bicluster be merged into a new nER sub-bicluster. This new nER sub-bicluster comprises the n exchange rates in the 2ER sub-bicluster and the (n-1)ER sub-bicluster. The time point set contained in this new nER sub-bicluster is the intersection of the time point sets of the 2ER sub-bicluster and the (n-1)ER sub-bicluster. The type of this new nER sub-bicluster is the same as those of the two original sub-biclusters.

We put all the 2ER sub-biclusters detected in the first phase of our biclustering algorithm into set Ω. Set Ψ is used to contain the generated complete biclusters in the process of the expansion algorithm. The expansion algorithm consists of the following steps:

(1) Remove the 2ER sub-biclusters that can not be classified into either one of the two bicluster types that we concern from Ω.

(2) Select the first 2ER sub-bicluster b_1 in Ω as a temporary complete bicluster t and remove b_1 from Ω.

(3) Search through all the 2ER sub-biclusters in Ω and find the first 2ER sub-bicluster b that can be merged with the temporary complete bicluster t. Merge b and t

to generate the new temporary complete bicluster. Remove b from Ω.

(4) Repeat Step3 until no 2ER sub-bicluster in Ω can be merged with the temporary complete bicluster t. Here t will represent a final complete bicluster. Put t into Ψ. Search through all the 2ER sub-biclusters in Ω and remove all the ones that can be regarded as part of t.

(5) Repeat Step 2 to Step 4 until Ω is empty. Here all the complete biclusters are recorded in Ψ.

4. Experimental result

We use an exchange rate dataset to test the performance of our biclustering algorithm. All the historical exchange rate data are obtained from the internet. Seventeen currencies are selected for our experiments. They are ARS (Argentine Peso), AUD (Australian Dollar), BRL(Brazilian Real), CAD (Canadian Dollar), CHF (Swiss Franc), EGP (Egyptian Pound), GBP (British Pound), IDR (Indonesian Rupiah), INR (Indian Rupee), JPY (Japanese Yen), MXN (Mexican Peso), PHP (Philippine Peso), RUB (Russian Rouble), SGD (Singapore Dollar), THB (Thai Baht), TWD (Taiwan Dollar), and ZAR (South African Rand). The exchange rates of these 17 currencies are all valued against the USD (US Dollar). The date range is from January 1, 1996 to December 31, 2005. a total of ten years. However, there is nearly no change in the exchange rates if measured day by day. This does not match our purpose of analyzing the principle of evolution of the exchange rate. Therefore, the comparison is made on a month by month basis. The average exchange rate for a month is used to represent the exchange rate of that month. Totally there are 120 months, so the size of the dataset is 120 by 17. Because the evolution of the exchange rate is what we are interested in, a difference dataset is generated from the original exchange rate dataset. Every element of this difference dataset is the difference between the exchange rate in a month and that lasting the preceding month. A positive element means an increase, a negative element means a decrease, and zero means no change. The size of this difference dataset becomes 119 by 17.

4.1. Results on simplified difference dataset

In this experiment we work on a simplified difference dataset. A threshold α is defined to make the simplification more reasonable. We use the symbol a to represent an element of the difference dataset. The simplification is as follows:

(1) If $a>0$ and $|a|>\alpha$, a is replaced with 1.

(2) If $a<0$ and $|a|>\alpha$, a is replaced with -1.

(3) If $|a|\leq\alpha$, a is replaced with 0

After simplification, the new dataset contains 1, -1 and 0 only. 1 means upward change, -1 means downward change and 0 means no change. The following is the analysis on two detected biclusters.

Bicluster1:

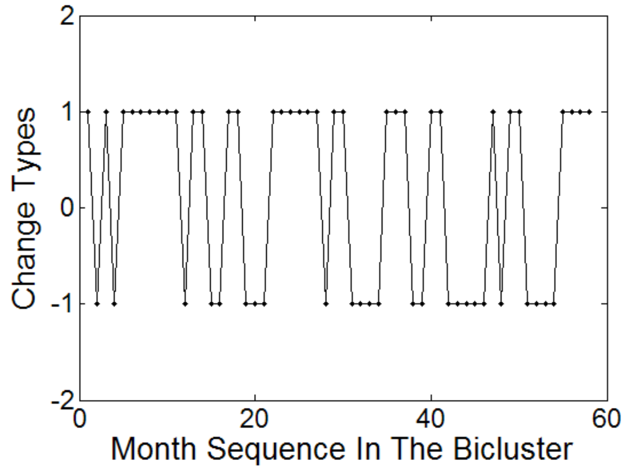

Figure 2. The common change types that the four exchange rates have in the 58 months.

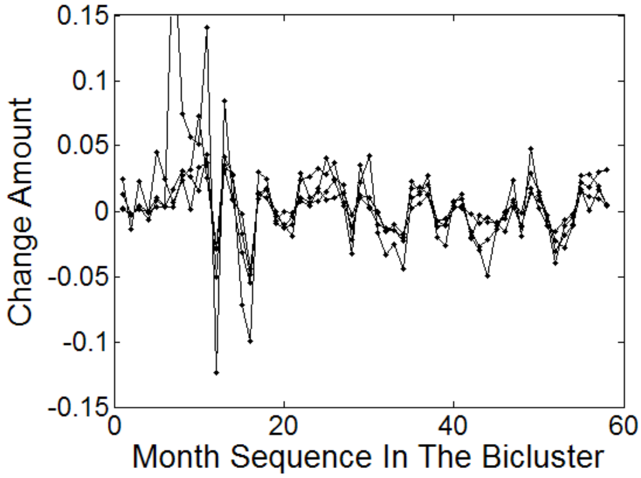

Figure 3. The corresponding real amount of proportional changes.

This bicluster is a constant row bicluster, which contains four exchange rates and 58 months. Figure 2 shows the common change types that the four exchange rates have in the 58 months, and Figure 3 presents the corresponding real amount of proportional changes. The four currencies contained in this bicluster are JPY, SGD, THB and TWD. The change types in the exchange rates for these four currencies are the same in 58 months. These discontinuous 58 months nearly equals five years, which is half of ten years. Therefore, this bicluster reflects that JPY, SGD, THB and TWD have a very close relationship. The conclusion accords with the fact in the real world. Japan, Singapore, Thailand and Taiwan are all in Far East. They are geographically very close to each other. Their economies are also related with each other very closely. Japan is the second strongest developed country in the world, and it is the most developed country in Asia. The other three countries or regions were all developing fast. There is plenty of cooperation between these four economies. Japan's financial policy has a certain degree of impact on the financial policies of all other countries or regions. Therefore this detected bicluster accurately reflects the close relationship between the four economies.

Bicluster 2:

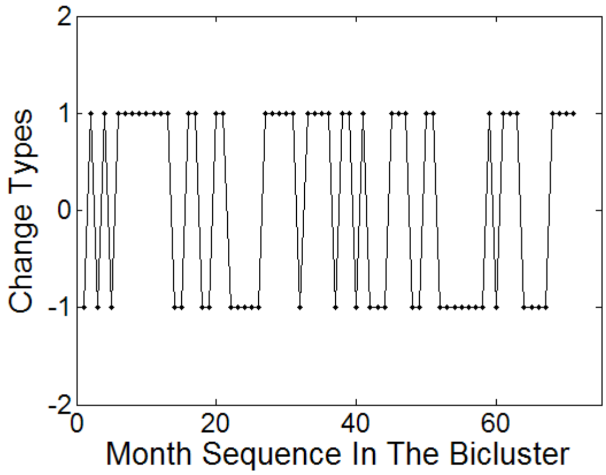

Figure 4. The common change types that the three exchange rates have in the 71 months.

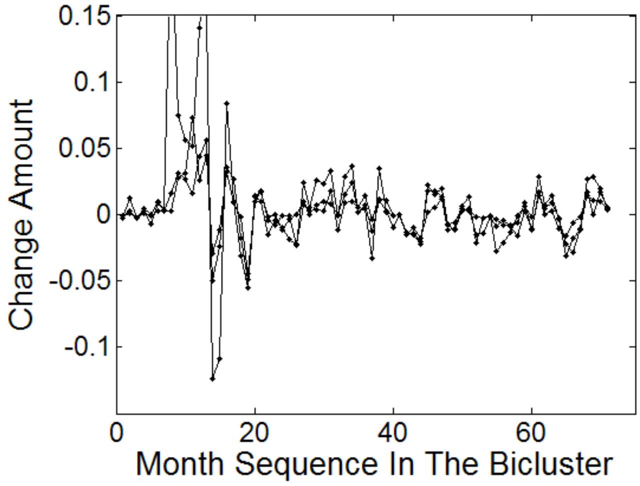

Figure 5. The corresponding real amount of proportional changes.

This bicluster is also a constant row bicluster, which contains three exchange rates and 71 months. Figure 4 shows the common change types that the three exchange rates have in the 71 months, and Figure 5 presents the corresponding real amount of proportional changes. The three currencies contained in this bicluster are SGD, THB, and TWD. The change types in the exchange rates for these three currencies are the same in 71 months, which are not continuous and occupy sixty percent of the total 119 months. Just as the analysis for Bicluster 1, this bicluster accurately reflects the close relationship in the economies of the three regions, Singapore, Thailand and Taiwan. However, if we observe the bicluster carefully, we see that the changes in the exchange rates for the three currencies are all up from the 8th month to the 13th month in the bicluster, and the increment is larger than the usual level. The period from the 8th month to the 13th month in the bicluster corresponds to the period from July 1997 to January 1998 without September 1997. What happened in that period? The answer is the finance crisis of Asia in 1990s. This crisis begins from July of 1997, and lasted almost two years. At the starting phase of this crisis, the exchange rates for THB, SGD and TWD increased quickly. After nearly half a year, the exchange rates for these currencies become steady. Therefore, this detected bicluster is significant and reflects the happenings in the real currency exchange world. The experiment proves the usefulness of the biclustering based technique proposed in this paper.

4.2. Results on real difference dataset

In this experiment, the real difference dataset is used directly to test the performance of the proposed biclustering algorithm. The following is the analysis on a detected bicluster.

Bicluster 1:

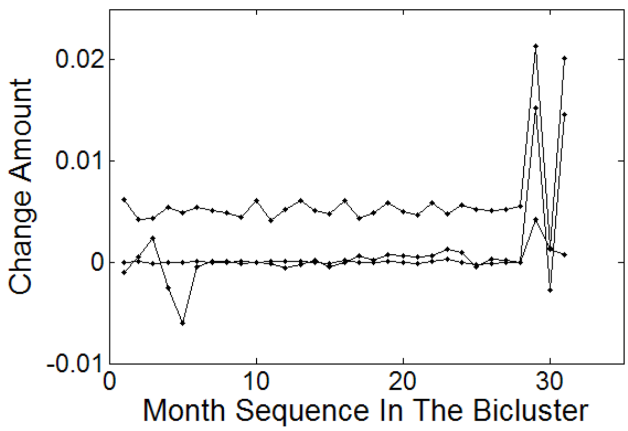

Figure 6. The real amount of proportional changes in the exchange rates for the three currencies in the 31 non-continuous months.

This bicluster contains three exchange rates and 31 months. The three currencies contained in this bicluster are ARS, BRL, and EGP. Figure 6 presents the real amount of proportional changes that the three exchange rates have in the 31 months. This bicluster belongs to the additive model, though the amounts of change in the exchange rate for the three currencies do not fulfill the additive relationship accurately in the last four months. The first 27 months of the total 31 months are all in the period from February 1996 to July 1998. In addition, there are 30 months from February 1996 to July 1998. That is to say, almost all of the months from February 1996 to July 1998 are contained in this bicluster. In this period, the amount of change in the exchange rate for BRL can be regarded as approximately 0.5 percent, and the amount of change in the exchange rates for ARS and EGP can be regarded as approximately 0. These are all the essential information contained in this bicluster. Now, how about the fact that the exchange rates for the three currencies in the period from February 1996 to July 1998? In this period, Argentina and Egypt both implemented a strictly fixed exchange rate regime, which means that there is nearly no change in the exchange rate. Brazil also implemented a fixed exchange rate regime, but it made the exchange rate increase by 6 percent each year by design. Six percent in one year means 0.5 percent in one month. We find that the essential information contained in this bicluster accurately accord with the fact.

5. Conclusion

In this paper, we have presented a biclustering based algorithm to analyze the correlation between different currency exchange rates. We use the FHT to detect sub-biclusters in the exchange rate pair spaces, and then use the expansion algorithm to combine the sub-biclusters to form the complete biclusters. We generate several exchange rate datasets to verify the performance of our method. Experiment results show that our method is effective. The bicluster patterns are consistent with the underlying economic reasons.

References

[1] Helmut Herwartz, and Henning Weber. “Exchange rate uncertainty and trade growth－a comparison of linear and non-linear (forecasting) models”, *Applied Stochastic Models in Business And Industry*, Vol.21, No.1, pp.1-26, 2005.

[2] Ansgar Belke, and Matthias Göcke. “Exchange rate uncertainty and employment: an algorithm describing ‘play’”, *Applied Stochastic Models in Business and Industry*, Vol.17, No.2, pp.181-204, 2001.

[3] Helmut Herwartz, and Hans-Eggert Reimers. “Empirical modeling of the DEM/USD and DEM/JPY foreign exchange rate: Structural shifts in GARCH-models and their implications”, *Applied Stochastic Models in Business and Industry*, Vol.18, No.1, pp.3-22, 2002.

[4] Shinji Takagi. “The Yen and Its East Asian Neighbors, 1980-95: Cooperation or Competition?”, *NBER Working Paper No.5720*.

[5] Laura Lazzeroni, and Art Owen. “Plaid models for gene expression data”, *Statistica Sinica*, Vol.12, pp.61-86, 2002.

[6] Yizong Cheng, and George M. Church. “Biclustering of expression data”, In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB)*, Vol.8, pp.93-103, 2000.

[7] Inderjit S. Dhillon. “Co-clustering documents and words using bipartite spectral graph partitioning”, In *Proceedings of the 7th ACM SIGKDD,* pp.269-274, 2001.

[8] Xiangchao Gan, AWC Liew, and Hong Yan. “Biclustering gene expression data based on high dimensional geometric method”, In *Proceedings of 4th International Conference on Machine Learning and Cybernetics,* Guangzhou, Vol.6, pp.3388-3393, 2005.

[9] H Li, MA Lavin, and RJ Le Master. “Fast Hough transform: a hierarchical approach”, *Computer Vision, Graphics, and Image Processing,* Vol.36, pp.139-161, 1986.

[10] Hongya Zhao, AWC Liew, and Hong Yan. “A new strategy of geometrical biclustering for microarray data analysis”, *In Proceedings of the 5th Asia-Pacific Conference on Bioinformatics Conference*, Imperial College Press, pp.47-56, 2007