Joel Chacón Castillo - Proyecto de verano

1 Análisis de algoritmos multiobjetivos

1.1 Representación ilustrativa de soluciones no dominadas

La optimización multi-objetivo trata con una multiple información. Graficando la función objetivo de la mejor solución en una iteración el rendimiento de un algoritmo puede ser demostrada. Sin embargo en optimización multi-objetivo existe más qu eun objetivo y en muchos casos de interés se comportan de una forma conflictiva. Se acostumbra mostrar las solucones no dominadas en un gráfico o plot del espacio de funciones objetivo, las variables de decisión de las soluciones no dominadas pueden ser mostradas por una técnica de ilustración.

Scatter-Plot Matrix Method

Realiza el gráfico de todos las combinaciones $\binom{M}{2}$ de plots en las M funciones objetivo.

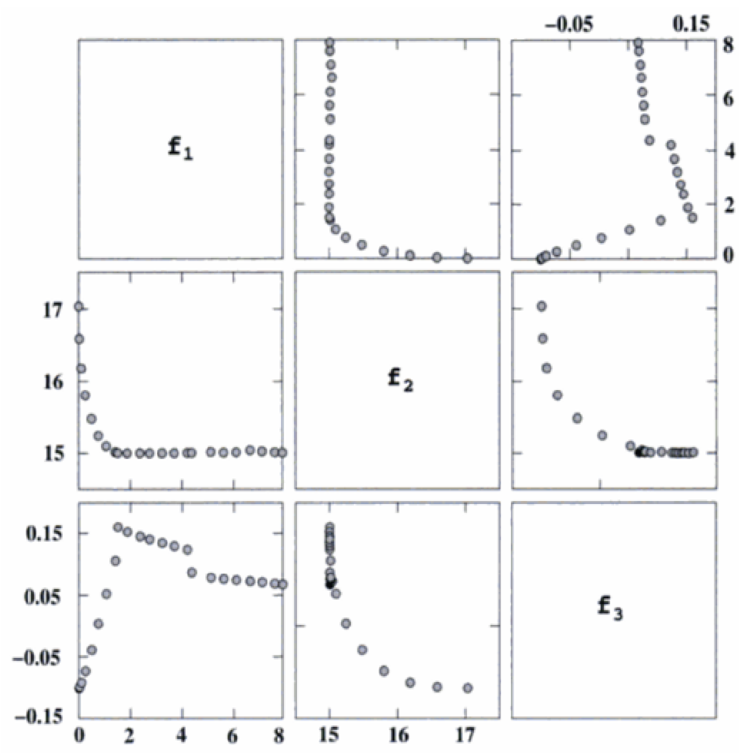

Figure 1: El método "Scatter-Plot Matrix".

Value Path Method

Esta es una popular forma para mostrar las soluciones no dominadas obtenidas para un problema el cual está conformado por más de dos objetivos. Los ejes horizontales marcan la identidad de la función objetivo y entonces debe ser establecido sólo en enteros iniciando desde 1. En caso de existir M objetivos, entonces deberían existir M marcas en este eje. El eje vertical representa los valores de la función objetivo normalizados. Cuando todas las soluciones del conjunto no dominado obtenido son graficadas en esta forma, el gráfico presenta cierta información:

- Para cada función objetivo, los valores de la función extrema proveen una cualitativa aignación de la dispersión en las soluciones obtenidas. Un algoritmo el cual tiene sus soluciones dispersas sobre la barra entera es considerada a ser buena en encontrar soluciones diversas.
- El grado en que las líneas cruzadas "zig-zag" muestran las posibilidades entre las funciones objetivo capturadas por las soluciones no dominadas obtenidas. Un algoritmo el cual tiene un largo cambio de la pendiente entre dos barras de funciones objetivo consecutivas sin consideradas como buenas en términos de encontrar buenas soluciones no dominadas.

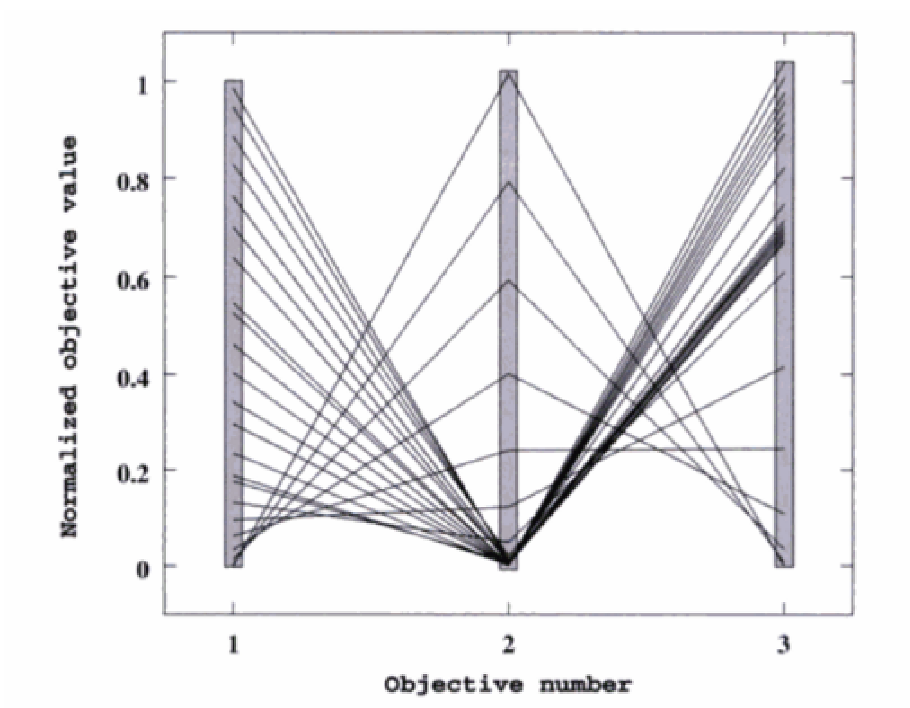

Figure 2: El método "Value path".

Bar Chart Method

Otra forma de representar soluciones no dominadas es graficando las soluciones como una gráfica de barras. Primeramente las soluciones no dominadas son ordenadas en un orden particular. Posteriormente para cada función objetivo, los valores de cada solución en el mismo orden son graficados con una barra. Desde que los objetivos pueden tomar distintos rangos de valores, se acostumbra graficar las barras con valores de los objetivos normalizados. De esta forma si existen N soluciones obtenidas, N barras diferentes son graficadas por cada función objetivo. Dado que las barras son graficadas la diversidad en distintas soluciones para cada objetivo pueden ser directamente observados del plot.

Star Coordinate Method

Un sistema coordinado de estrella representa múltiples soluciones no dominadas. Para M funciones objetivo, un círculo es dividido en M arcos iguales. Cada linea radial conecta el fin de un arco con el centro del círculo los cuales representan los ejes función objetivo. Desde que el rango de cada función objetivo puede ser diferentees necesario etiquetas el final de cada línea. La normalización de los objetivos no es necesario. Una vez que el framework esté listo, cada solución puede ser marcada en un círculo. Entonces para N soluciones, existirán N círculos. Visualmente éstos transmitirán la convergencia y diversidad en las soluciones obtenidas.

1.2 Métricas para medir la aproximidad al frente de pareto óptimo

Error Ratio

Esta métrica simplemente contabiliza el número de soluciones del conjunto $\mathbf{Q}$ los cuales no son miembros del conjunto de pareto óptimo $\mathbf{P}^*$

$$ER = \frac{\sum_{i=1}^{|Q|} e_i}{|Q|} \tag{1}$$

Donde $e_i = 1$ si $i \notin \mathbf{P}^*$ de otra forma $e_i = 0$, un menor valor de ER significa una mejor convergencia de frente de pareto óptimo. Los valores que toma ER es entre cero y uno, $ER = 0$ significa que todos los miembros son soluciones de frente de pareto óptimo $\mathbf{P}^*$, y $ER = 1$ significa que ningún miembro es solución de $\mathbf{P}^*$. La desventaja de este método es que si ningún elemento de Q pertenece al frente óptimo, entonces no se tiene una medición de la proximidad, esta métrica puede ser más útil redefiniendo e_i, para cada solutión $i \in Q$, si la distancia euclidiana mínima (en el espacio objetivo) entre i y $\mathbf{P}^*$ es mayor que el límite δ entonces el parámetro e_i es asignado como la unidad.

Set Coverage Metric

Esta métrica es similar a Error Ratio, pero también puede ser utilizada para obtener una idea de la dispersion relativa en dos conjuntos de vectores solución A y B. $\zeta(A,B)$ realiza el cálculo de la proporción de soluciones en B los cuales son dominados por las soluciones de A:

$$\zeta(A,B) = \frac{|\{b \in B | \exists a \in A : a \preceq b\}}{|B|} \tag{2}$$

$\zeta(A,B) = 1$ indica que todos los miembros de B son débilmente dominados por A, $\zeta(A,B) = 0$ indica que ningún miembro de B es débilmente dominado por A. Esta métrica también se utiliza para comparar el rendimiento de dos algoritmos, opcionalmente dada la solución del frente óptimo no dominado $\mathbf{P}^*$ la métrica $\zeta(\mathbf{P}^*, Q)$ determina la proporción de soluciones en Q las cuales son débilmente dominadas por los miembros de $\mathbf{P}^*$, se desea que $\zeta(Q, \mathbf{P}^*)$ sea siempre cero.

Generational Distance

Esta métrica encuentra una distancia promedio de las soluciones desde Q a $\mathbf{P}^*$ de la siguiente forma:

$$GD = \frac{(\sum_{i=1}^{|Q|} d_i^p)^{1/p}}{|Q|} \tag{3}$$

Para $p = 2$ el parámetro d_i es la distancia euclideana entre la solución $i \in Q$ y el miembro más cercano de $\mathbf{P}^*$:

$$d_i = min_{k=1}^{|\mathbf{P}^*|} \sqrt{\sum_{m=1}^{M} (f_m^{(i)} - f_m^{*(k)})^2} \tag{4}$$

donde $f_m^{*(k)}$ es el m-ésima valor de la función objetivo del k-ésimo miembro de $\mathbf{P}^*$. La dificultad con esta métrica s que si existe un Q para la cual hay una larga fluctuación en los valores de las distancias, esta métrica podría no revelar la distancia verdadera, en tal caso se debe efectuar el calculo de la varianza de la métrica GD. Además si los valores de las funciones objetivo son de distinta magnitud, éstas deberían ser normalizadas antes de implementar la métrica. Esta métrica es identificada por el símbolo γ y si la desviación estándar de γ es un valor pequeño el cálculo de γ es aceptado.

1.3 Métricas para evaluar la diversidad entre soluciones no dominadas

Spacing

Esta métrica es calculada con la medida de la distancia relativa entre soluciones consecutivas en el conjunto no dominado de la forma:

$$S = \sqrt{\frac{1}{|Q|} \sum_{i=1}^{|Q|} (d_i - \overline{d})^2} \tag{5}$$

donde $d_i = min_{k \in Q \wedge k \neq i} \sum_{m=1}^{M} |f_m^i - f_m^k|$ y $\overline{d}$ es el valor promedio de las distancias $\overline{d} = \sum_{i=1}^{|Q|} d_i/|Q|$. La distancia es el valor mínimo de la suma de las diferencias absolutas en los valores de la función objetivo entre la i-ésima solución y cualquier otra solución del conjunto no dominado. Esta métrica mide la desviación estándar de los valores distinto de d_i. Cuando las soluciones están cercanas y uniformemente espaciadas, la medición de la distancia será pequeña. Esta métrica no toma en cuenta la calidad de la disperción en las soluciones, es decir, siempre que la dispersión es uniforme en un rango de soluciones obtenidas, la métrica S produce un valor pequeño.

Spread

Esta métrica toma en cuenta la calidad de la dispersión en las soluciones obtenidas

$$\triangle = \frac{\sum_{m=1}^{M} d_m^e + \sum_{i=1}^{|Q|} |d_i - \overline{d}|}{\sum_{m=1}^{M} d_m^e + |Q|\overline{d}} \tag{6}$$

donde d_i puede ser cualquier distancia en las soluciones vecinas y $\overline{d}$ es la distancia promedio. La distancia euclidiana, la suma de las diferencias absolutas en los valores objetivo o en la distancia de crowding pueden ser utilizados para calcular d_i. El parámetro d_m^e es la distancia entre las soluciones extremas de $\mathbf{P}^*$ y Q correspondiendo a la m-ésima función objetivo. Esta méttrica toma el valor de cero en una distribución ideal, sólo cuando $d_j^e = 0$ y todos los valores d_i son idénticos a la media $\overline{d}$, la primer condición significa que en el conjunto solución no dominado la solución verdadera existe y la segunda condición indica que la distribución de soluciones intermedias es uniforme. Entonces para una distribución ideal de soluciones no dominadas $\triangle = 0$.

Maximum Spread

Define una métrica midiendo la longitud de la diagonal de una hipercaja "hyperbox" formada por los extremos de los valores observados del conjunto no dominado observado.

$$D = \sqrt{\sum_{m=1}^{M} (|Q|_{maxf_m^i} - |Q|_{minf_m^i})^2} \tag{7}$$

Para problemas de dos funciones objetivo, esta métrica se refiere a la distancia euclidiana entre las soluciones extremas del espacio objetivo. La versión normalizada es:

$$\overline{D} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(\frac{max|Q|f_m^i - min|Q|f_m^i}{F_m^{max} - F_m^{min}} \right)^2} \tag{8}$$

donde $F_m^{max} y F_m^{min}$ son los valores máximos y minimos del m-ésimo objetivo en el conjunto óptimo de pareto $\mathbf{P}^*$. Si la métrica es uno, se tiene un conjunto ampliamente disperso de soluciones.

Chi-Square-Like Deviation Measure

Esta métrica se utiliza para funciones multi-modales, en esta métrica es utilizado un parámetro de vecindario ϵ para contar el número de soluciones η_i en cada solución de pareto óptima (solución $i \in \mathbf{P}^*$). El cálculo de la distancia puede efectuarse en el espacio de las variables de decisión o en el espacio de función objetivo. La desviación entre la cantidad de este conjunto de números con un conjunto ideal es medido con sentido de "chi-square":

$$\imath = \sqrt{\sum_{i=1}^{|\mathbf{P}^*|+1} (\frac{\eta_i - \overline{\eta_i}}{\sigma_i})^2} \tag{9}$$

Dado que no existe una solución de pareto particular es utiliza la distribución uniforma como distribución ideal. Esto indica que deben existir $\overline{\eta_i} = |Q|/|\mathbf{P}^*|$ soluciones en el nicho de cada solución óptima de pareto. El parámetro $\sigma_i^2 = \overline{\eta}_i(1-\overline{\eta}_i/|Q|)$ es sugerido para $i = 1, 2..., |\mathbf{P}^*|$. Sin embargo el índice $i = |\mathbf{P}^*|+1$ representa todas las soluciones las cuales no residen en el vecindario ϵ para cualquier solución escogida del frente óptimo de pareto. Para este índice el número ideal de soluciones y su varianza son calculados de la forma:

$$\begin{aligned} \overline{\eta}_{|\mathbf{P}^*|+1} &= 0 \\ \sigma^2_{|\mathbf{P}^*|+1} &= \sum_{i=1}^{|\mathbf{p}|^*} \sigma_i^2 = |Q|(1 - \frac{1}{|\mathbf{P}^*|}) \end{aligned} \tag{10}$$

Si la distribución ideal es encontrada esta métrica tendrá un valor de cero. Si muchas soluciones estás lejos del frente óptimo de pareto, esta métrica no puede evaluar la dispersión de las soluciones adecuadamente.

1.4 Metricas que evlúan cercanía al frente de pareto óptimo y diversidad de la solución no dominada

Existen varias métricas que pueden ofreces una medición cualitativa de convergencia y de diversidad. Sin embargo pueden ser utilizadas otra métrica para obtener una mejor evaluación.

Hypervolume

Esta métrica calcula el volúmen (en el espacio objetivo) cubierto por miembros de Q (el conjunto solución calculado) para problemas donde todos los objetivos se van a minimizar, es decir para cada solución $i \in Q$, un hipercubo v_i es construido con un punto de referencia W y la solución i como las esquinas diagonales del hipercubo. El punto de referencia puede ser encontrado construyendo el vector de los peores valores de las funciones objetivo. Posteriormente la unión de todos los hipercubos es encontrada y su hipervolúmen (HV) es calculado:

$$HV = volume(\cup_{i=1}^{|Q|} v_i) \tag{11}$$

Esta metrica no es libre para objetivos arbitrariamente escalados. por lo que esta métrica puede ser evaluada usando funciones objetivo nomalizados.
Otra forma de eliminar el sesgo en cierta medida y calcular un valor normalizado es utilizando la métrica HVR el cual es la razón del HV de Q y de $\mathbf{P}^*$:

$$HVR = \frac{HV(Q)}{HV(\mathbf{P}^*)} \tag{12}$$

Para un problema donde todos los objetivos se van a minimizar, el mejor (máximo) valor de HVR es uno (cuando $Q = \mathbf{P}^*$).

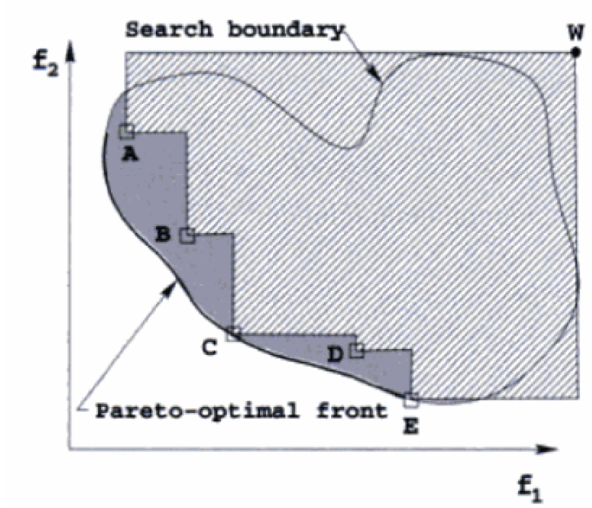

Figure 3: El hipervolúmen encerrado por soluciones no dominadas.

Attaiment Surface Based Statistical Metric

Establece un envolvente el cual puede ser formado marcando todas las soluciones en el espacio de búsqueda que seguramente son dominados por un conjunto de soluciones no dominadas. El envolvente generado es llamado "attainment surface" y es idéntica a la superficie utilizada para calcular el hipervolúmen. En la práctica un MOEA se ejecutará múltiples veces, cada vez iniciando el MOEA de una población inicial distinta o configuración deparámetros distintos. Posteriormente las soluciones no dominadas pueden ser utilizadas para encontrar un "attainment surface" para cada ejecución, se espera que existan variaciones en el los resultados.Ćuando dos o más algoritmos son comparados el clustering de las soluciones cercanas al frente óptimo de pareto podría no proveer una idea clara acerca de que algoritmo tiene un mejor rendimiento. En primer lugar una serie de líneas diagonales imaginarias, corriendo en la direción enque todos los objetivos mejoras son escogidos. Para cada línea los puntos que intersectas de todos los "attainment surface" para un algoritmo son calculados. Estos puntos

estarán en la línea escogida y entonces tendrán una frecuencia de distribución. La distribución de frecuencia y distintos "attainment surfaces" para otro conjunto de soluciones no dominadas utilizando un segundo algoritmo pueden ser calculados igualmente. Una vez que las distribución de frecuencias son encontradas en una línea escogida se implementa un test estadístico o el test de Kolmogorov-Smirnov se puede utilizar con un nivel de confianza β para concluir cual algoritmo es mejor a lo largo de la línea. Este mismo procedimiento puede ser repetido para otras líneas que cruzan los puntos. En cada línea una de las tres decisiones pueden ser efectuadas con un nivel de confianza escogido: (i) El algoritmo A es mejor que el algoritmo B, (ii) El algoritmo B es mejor que el algoritmo A, o (iii) no hay una conclusión con el nivel de confianza establecido. Con un totald e L líneas elegidas, se sugiere una métrica $[a, b]$ con un límite de confianza β, donde a es el porcentaje de veces que el algoritmo A fue mejor y b es el porcentaje de las veces en que el algoritmo B fue mejor. Entonces $100 - (a + b)$ proporciona el procentaje de veces en que el resultado fue estadísticamente no concluso. Se puede concluir que el algoritmo A es mejor que el algoritmo B si $a > b$ o viceversa. Se menciona que los resultados dependen del nivel de confianza establecido.
Este procedimiento se puede extender a más de dos algoritmos para k algoritmos se efectúa el proceso:

- El procentaje (a_k) de una región donde uno puede ser estadísticamente confiable con el nivel de confianza escogido para el algoritmo k no fué "vencido" por otro algoritmo.

- El porcentaje (b_k) de la región donde uno puede ser estadísticamente confiable con el nivel de confianza escogido que el algoritmo "vence" a todos los otros $(k-1)$ algoritmos. Un algoritmo con valores largos de a_k y b_k es considerado bueno. Dos o más algoritmos teniendo mas o menos los mismos valores de a_k debe ser juzgado por b_k. En tal caso el algoritmo que tiene valores más largos de b_k es mejor.

Weighted Metric

Un procedimiento simple para evaluar los dos objetivos (obtener diversidad y converegencia al frente óptimo de pareto), puede ser definir una métrica de pesos combinando una de las métricas de convergencia y una de las métricas de diversidad:

$$W = w_1 GD + w_2 \Delta \tag{13}$$

con $w_1 + w_2 = 1$. Aquí se ha combinado la métrics de distancia generacional (GD) para evaluar la convergencia y Δ mide la preservación de la diversidad del algoritmo. Un algoritmo que tiene un valor pequeño de W significa que el algoritmo es bueno en ambos aspectos. El usuario puede utilizar los pesos apropiados w_1 y w_2 para combinar las dos métricas, sin embargo si esta métrica se utiliza es mejor que emplear un par de de métricas normalizadas.

Non-Dominated Evaluation Metric

Ya que ambas métricas se evalúan como objetivos en conflicto, es ideal plantear la evaluación de MOEAs un problema de dos objetivos. Si los valores de la métrica para un algoritmo domina al otro algoritmo, entonces el primero es sin duda mejor que el último. De otra forma no se puede afirmar ningúna conclusión acerca de los algoritmos..