Abstract:

This paper introduces a new set of descriptors for usage in cheminformatics.

A gradient of Electrostatic [Potential] Energy, or GEE, Descriptors. Consider a molecule so either Cartesian or spherical coordinates can be used to describe the atoms present in the compound:

Electrostatic potential for an atom directly correlates with or is proportional to the atom’s electronegativity:

\[ V_{E,i} \propto \chi_i \]
where \( V_{E,i} \) is the electrostatic potential of atom \( E \) and \( \chi \) is the electronegativity of atom \( i \). Thus, the electrostatic potential energy for atom \( i \), or \( E_i \), can be expressed as the following:

\[
E_i \propto \chi_i \rho_i
\]

where \( \rho_i \) is the magnitude of the distance of atom \( i \) energy from an origin, or the center of the molecule. The above would suggest the total electrostatic energy of a molecule about its center would be defined as:

\[
E = \sum_i \chi_i \rho_i
\]

The gradient of electrostatic potential energy about the center of a molecule would be as follows:

\[
\nabla E = \nabla \sum_i \chi_i \rho_i
\]

where the \( \nabla \) is the del operator. After the application of the chain rule of differentiation, one obtains:

\[
\nabla E = \nabla \sum_i \chi_i \rho_i
= \sum_i \rho_i \nabla \chi_i + \sum_i \chi_i \nabla \rho_i
\]

Since the electronegativity of an atom is relatively constant, the left term of the above expression drops out leaving:

\[
\nabla E = \sum \chi_i \nabla \rho_i
\]

Assuming one is working with spherical coordinates, the gradient of \( \rho_i \) will reduce to:

\[
\nabla \rho_i = \frac{1}{\rho_i} \langle x_i, y_i, z_i \rangle,
\]

where \( x_i, y_i, \) and \( z_i \) are the Cartesian coordinates of atom \( i \) from the center of the molecule.

Substituting in the spherical coordinates associated with the Cartesian system:

\[
x = \rho \cos \theta \sin \phi,
\]

\[
y = \rho \sin \theta \sin \phi,
\]

\[
z = \rho \cos \phi.
\]

The gradient of \( \rho_i \) simplifies to:

\[
\nabla \rho_i = \langle \cos \theta_i \sin \phi_i, \sin \theta_i \sin \phi_i, \cos \phi_i \rangle.
\]

Knowing \( \theta \) and \( \phi \) are defined as:
then the gradient of $\rho_i$ becomes:

$$\nabla \rho_i = \left( \cos \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right), \sin \left( \arctan \left( \frac{z_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right), \cos \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right) \right).$$

Thus, the gradient of the electrostatic potential energy about the center of the molecule is given as the following expression:

$$\nabla E = \sum_i \chi_i \left( \cos \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right), \sin \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right), \cos \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right) \right).$$

Next, one can define a set of descriptors based upon the distance from the center of a compound:

$$GEE \left( \rho \right) = \sum_i \sum_j \left( u_{\rho_j} \left( \rho \right) - u_{\rho_j+40} \left( \rho \right) \right) \chi_i \left( \cos \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right), \sin \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right), \cos \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right) \right).$$

which can be reduced to Cartesian components:

$$GEE_{j,x} \left( \rho \right) = \sum_i \sum_j \left( u_{\rho_j} \left( \rho \right) - u_{\rho_j+40} \left( \rho \right) \right) \chi_i \cos \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right),$$

$$GEE_{j,y} \left( \rho \right) = \sum_i \sum_j \left( u_{\rho_j} \left( \rho \right) - u_{\rho_j+40} \left( \rho \right) \right) \chi_i \sin \left( \arctan \left( \frac{y_i}{x_i} \right) \right) \sin \left( \arccos \left( \frac{z_i}{\rho_i} \right) \right),$$

$$GEE_{j,z} \left( \rho \right) = \sum_i \sum_j \left( u_{\rho_j} \left( \rho \right) - u_{\rho_j+40} \left( \rho \right) \right) \chi_i \frac{z_i}{\rho_i}.$$

Finally, the net GEE descriptors for $\rho$ at $j$ would be defined as the following expression:

$$GEE_1 \left( \rho \right) = \sqrt{GEE_{j,x}^2 \left( \rho \right) + GEE_{j,y}^2 \left( \rho \right) + GEE_{j,z}^2 \left( \rho \right)}.$$