A Proposed General Formula to Create and Analyze Baking Recipes

Michael Ohene

EasierBaking,
1301 St. John Street, Apt. 206, Lafayette, LA 70506
michael@easierbaking.com
http://easierbaking.com

Abstract. A mathematical formula for characterizing baking recipes is presented as part of the 2017 Computer Cooking Contest Open Challenge. The formula produces three characteristic values, which along with common knowledge rules and classification, form the basis of two computer applications: Random Recipe Generator, which creates recipes, and Recipe Report Card, which analyzes recipes.

Keywords: recipe analysis, recipe creation, recipe classification

1 Introduction

The mystery of baking recipes has existed for many years despite many attempts to discover a formula or set of rules to describe them [1], [2]. The discovery of a universal formula or set of rules would, at least, form a basis for answering key questions governing baking. Of particular interests are the abilities to create custom recipes and to discover new uses for ingredients in baking. In lieu of a universal formula, creating new recipes by adaptation remains popular, however, this approach results in recipes limited by their reference recipe.

Adaptation has been formalized in research communities, where it involves creating new recipes by the introduction of substitute ingredients [6], primarily in a like-for-like relationship, and adaptation rules. The methods for substituting ingredients have involved evaluating the validity of substitutions by a scoring procedure [3] and by ingredient generalization through a cooking ontology [7].

This paper outlines an extended, generalized substitution process where any ingredient is a candidate for substitution. The only restrictions are common knowledge rules placed on baked good recipes (e.g., "Cobbler must not contain water", "Pie crust must contain water"). To avoid the tedious work alluded to in [7], the scope of this procedure shall be limited to baked goods.

1.1 What is a Baking Recipe?

A baking recipe provides a list of ingredients and measurements, which includes instructions for combining the ingredients. Each ingredient may be considered

either a wet ingredient, a dry ingredient, or semi-wet ingredient. In the following procedure, first detailed in [5], wet and semi-wet ingredients are given constant values (see Table 1), while flavorings, leavenings (e.g., baking powder, baking soda, yeast, etc.), seasonings (e.g., salt), and food pieces (e.g., shredded coconut, walnut pieces, sesame seeds, etc.) are ignored. The constant values are multiplied by their respective measurements (usually in cups) to yield a numerical product. The products are summed and finally divided by the dry ingredient product(s), obtained from values in Table 1, to yield solutions called the moistness, fat, and egg value [5]. These characteristic values (i.e., the moistness value, the fat value, and the egg value) complete the characterization of baked good recipes.

Ingredients	**Value per Cup**
Wet Ingredients	
Water/Juice/Water/Milk	1
Butter/Oil	0.50
Banana	0.375
*Large egg (50 grams)	0.167, 1
Honey/Molasses	0.70
Dry Ingredients	
Flour (all-purpose, cocoa powder, whole-wheat)	1
Old-fashioned rolled oats	0.50
Semi-wet Ingredients	
Ground nuts (almond, pecans, walnuts)	0.33

Table 1. Constants for common wet, dry, and semi-wet ingredients. The large egg constant does not use a per cup value. *Large eggs each have a value of 0.167 in the moistness calculation and 1 in the egg calculation. Constants for common dry ingredients.

Ingredients	**Measure**	**Wet Value**	**Dry Value**
All-purpose flour, Cups(g)	2 1/2, (352g)	—	2.50
Butter, Tbsp(g)	16, (224g)	0.50	—
Egg, #(g)	1, (50g)	0.167	—
Confectioner's sugar, Cups(g)	1 1/2, (120g)	—	—
Vanilla extract, tsp(g)	1, (4g)	—	—
Almond extract, tsp(g)	1/2, (2g)	—	—
Salt, tsp(g)	1/2, (3g)	—	—
Baking soda, tsp(g)	1, (5g)	—	—
Cream of Tartar, tsp(g)	1, (5g)	—	—

Table 2. Mary's Sugar Cookie recipe with moistness values. [12]

$$\frac{16Tbsp * \frac{1Cup}{16Tbsp} * \frac{1}{2} + 1 * \frac{1}{6}}{2.50 * 1} = 0.27 \quad (1)$$

Equation (1) shows the wet-over-dry ingredient equation used to calculate the moistness value from *Mary's Sugar Cookie* recipe in Table 2. A similar equation is used to calculate the fat value in equation (2), only using ingredients that are considered fats. The egg value requires a number-of-eggs-per-cup-of-dry-ingredients calculation shown in equation (3).

$$\frac{16Tbsp * \frac{1Cup}{16Tbsp} * \frac{1}{2}}{2.50 * 1} = 0.20 \quad (2)$$

$$\frac{1}{2.50 * 1} = 0.4 \quad (3)$$

The general linear equation

$$(\frac{1}{q_{n+1}i_{n+1}})(q_1i_1 + q_2i_2 + ...q_ni_n) = \left[\underline{y}, \overline{y}\right] \quad (4)$$

defines baked goods through the use of characteristic values, where i is the ingredient constant, q is the quantity, and n is the nth ingredient. The term $\left[\underline{y}, \overline{y}\right]$ refers to the numerical range in moistness, fat, or egg value of a baked good. $\underline{y}$ represents the lower limit and $\overline{y}$ represents the upper limit of the numerical range.

1.2 Knowledge Acquisition

To accurately define the numerical ranges corresponding to baked goods, the acceptability of recipes and recipe reviews were considered. Instead of analyzing the reliability of users as in [3], the sheer number of reviews and the selection of recipe-focused review sites - as opposed to blogger-focused review sites - served to minimize unreliable reviews. The recipe review ratings and the "make it again" ratings served to define "acceptability". From this point the acceptable linear equations were constructed from equation (4) to determine the unknown constants.

From the collection of recipes, acceptable recipes tended fall within the predefined numerical ranges, thereby satisfying equation (4). Unacceptable recipes tended to fall outside the predefined numerical ranges of the baked goods. Example deviations from these generalized numerical ranges for cakes are presented in bold text in Table 3. By *generalized*, it is meant that the numerical range used for cakes in Table 3 are aggregations of several independent numerical ranges representing a variety of cakes (e.g., the egg value for pound cake only occupies a portion of the 1.00-3.50 egg range).

Recipes	Characteristic Values			Comments, (Exceptions)
	Moistness	Fat	Egg	
Cakes	0.68-1.15	0.13-0.34	1.00-3.50	
Basic 1-2-3-4 Layer	0.72	0.17	1.3	
Devil's Food	0.83	0.18	1.1	
Glazed Lemon-Thyme	**1.86**	**0.67**	2.7	
Glazed Lemon-Thyme (corrected)	0.8	0.27	1	
Confetti	**1.29**	**0.38**	2.5	Possible bad recipe
Pineapple Curry	0.7	0.13	1	
Classic Pound	0.87	0.29	2.3	
Blueberry Cornmeal	1.05	0.16	1.3	
German Chocolate	**0.64**	**0.07**	**0.2**	Possible bad recipe
Nejla's Yogurt	**0.6**	**0**	1.5	(Sponge cake)
Italian Cream	0.75	0.21	1.6	
Champagne	**0.62**	0.18	**0.5**	(Wedding cake)
Tasted Just Like Wedding	**0.59**	0.17	**0.3**	(Wedding cake)
Angel Bean Food	**0.27**	**0**	1.6	(Angel food cake)
Old-Fashioned Coconut	0.81	0.17	1.3	
Peanut Butter and Chocolate Swirl	0.88	0.25	1.5	
Pecan Crumble	0.89	0.13	1	
Guinness Stout	1.00	**0.1**	**0.9**	Possible bad recipe
Vanilla Bean Angel Food	0.97	**0**	2	(Angel food cake)
Strawberry and Cream	0.81	0.25	1.3	
Caramel	0.72	0.17	1.3	
Meyer Lemon	0.98	0.21	1.1	
Old-Fashioned Red Velvet	1.01	0.2	1.6	
Spiced Crumb	0.96	0.17	1.3	
Blood Orange	**1.3**	0.17	2	
Blood Orange (corrected)	1.13	0.17	2	
Hummingbird	0.77	0.17	1	
Strawberry Buttermilk	1.01	0.13	1.3	
Tres Leches	1.06	0.2	1.3	
Rum-Soaked	0.86	0.22	3	
Upside Down Chocolate	0.97	0.2	1.5	
Cardamom Flourless	**1.60**	**0.6**	6	Possible bad recipe
Carrot	0.91	0.17	1.3	
Pear Almond	**0.52**	0.25	1.5	
Pear Almond (corrected)	**0.67**	0.25	1.5	Possible bad recipe

Table 3. Characteristic values from the 2016 *bakeFromScratch* Special Edition. The bold values are the values that fall outside the numerical range for (*cakes*). Some recipes, labeled (*corrected*), were corrected in the online edition of the magazine after receiving reader feedback. (*Moistness*) corresponds to the thinness of the batter. [10]

2 Random Recipe Generator

Random Recipe Generator uses characteristic values to provide users with unique, randomly generated recipes. The program simply converts characteristics values to recipes.

The Random Recipe Generator functions using a clickable photo grid of baked goods and two pull-down menus. The two pull-down menus allow users to choose their "Fat Level" and "Sweetness", by choosing between "Low Fat", "Regular Fat", or "High Fat" and "Not too Sweet", "Sweet", or "Really Sweet", respectively [4].

2.1 Choosing Characteristic Values

The steps for choosing a random recipe are as follows:
1) A click by the user selects the numerical ranges that define a baked good.
2) From the user's choice for fat level, the numerical range for fat, $\boldsymbol{x}_2$, is chosen.
3) Once the numerical range for fat, $\boldsymbol{x}_2 = [\underline{x}_2, \overline{x}_2]$, is chosen, a random fat value, x_2, is chosen and the other two values, the moistness value, x_1, and the egg value, x_3, are chosen according to the value x_2. Specifically,

a second value, x_3, in the numerical range for eggs, $\boldsymbol{x}_3 = [\underline{x}_3, \overline{x}_3]$, is randomly chosen, which in turn automatically sets the third value, x_1
or
a second value, x_3, in the numerical range for eggs, $\boldsymbol{x}_3 = [\underline{x}_3, \overline{x}_3]$, is randomly chosen, then a constant value is chosen such that a third value, x_1, lies within $[\underline{x}_1, \overline{x}_1]$.

2.2 Converting the Characteristic Values into a Recipe

After the process of choosing characteristic values based on the user input occurs, a base ingredient, i.e., an initial guess, is chosen, and the remaining ingredients are then substituted into equation (4). The possible measurements for the ingredients are defined by values in the Random Recipe Generator's database. In addition to measurement limits, the database also contains predefined, ingredient combinations. When equation (4)'s variables are replaced by quantities and ingredient constants, there exists some distance/error between the original random recipe's characteristic value vector, $\boldsymbol{x}$, and the substitution attempt's (adaptation's) characteristic value vector, $\boldsymbol{s_i}$, which can be calculated as the Euclidean distance, equation (5).

$$d(\boldsymbol{x}, \boldsymbol{s_i}) = \sqrt{(x_1 - s_{i,1})^2 + (x_2 - s_{i,2})^2 + (x_3 - s_{i,3})^2}. \tag{5}$$

There are 1410 iterations, i, of the ingredient substitution process, producing the distance values $d(\boldsymbol{x}, \boldsymbol{s_1}), ..., d(\boldsymbol{x}, \boldsymbol{s_{1410}})$. The ingredient substitution attempt (adaptation) with the shortest distance, $\arg\min d(\boldsymbol{x}, \boldsymbol{s_i})$, is selected and presented to the user.

2.3 Discovering New Ingredient Uses

Case-based reasoning differs from Random Recipe Generator's procedure, but it would be erroneous to say the current procedure did not utilize a case base. In fact, Random Recipe Generator relies upon a numerical abstraction of the recipe case base mentioned in the Knowledge Acquisition section. This abstraction helps to eliminate the detailed knowledge usually required to create recipes and completely eliminates the need for recipe retrieval.

In addition, instead of a more detailed formal concept analysis (FCA) approach described in [9], the only additional information needed to create a recipe is a generalized classification structure (e.g., whether an ingredient is a nut, egg, dairy, dry ingredient, chocolate, etc.). In other words, any ingredient can be added to the Random Recipe Generator database and incorporated into recipes as long as its classification and ingredient constant are known. As an example, Table 4 shows three recipes for chocolate chip cookies using peanut butter, ground almonds, and all-purpose wheat flour.

Ingredients	**Low Fat**		
	Not Too Sweet	**Sweet**	**Really Sweet**
All-purpose flour, Cups(g)	1 3/4, (247g)	1 1/2, (211g)	2 1/4, (317g)
Ground almonds, Cups(g)	1 3/4, (210g)	—	—
Peanut butter, Cups(g)	—	—	1/2, (129g)
Butter, Tbsp(g)	14, (196g)	8, (112g)	10, (140g)
Egg, #(g)	2, (100g)	1, (50g)	2, (100g)
Egg yolk, #(g)	—	—	—
Brown sugar, Cups(g)	2/3 Tbsp, (147g)	1/2, (110g)	1, (220g)
White sugar, Cups(g)	2/3, (133g)	1/2, (100g)	1, (200g)
Chocolate chips, Cups(g)	1 3/4, (319g)	1, (182g)	2, (365g)
Vanilla extract, tsp(g)	1 3/4, (8g)	1 1/4, (5g)	1 3/4, (8g)
Salt, tsp(g)	1/2, (3g)	1/4,(2g)	1/2, (3g)
Baking soda, tsp(g)	3/4, (4g)	1/2, (2g)	3/4, (4g)

Table 4. Three chocolate chip cookie recipes. (*Not Too Sweet*) , (*Sweet*), and (*Really Sweet*) correspond to low, normal, and high sweetness.

3 Recipe Report Card

A logical extension of the work in Table 4 is the development of a recipe analysis tool. In this role, Recipe Report Card serves to create an alternative to the traditional recipe review, i.e., to provide accurate, objective feedback for baking recipes. The use of the Recipe Report Card creates baking recipes which can be customized and prescreened. In addition, if the recipe's characteristic values fall within a predefined numerical range and satisfy common knowledge rules (e.g., "Brownie must contain chocolate"), the recipe is labeled and feedback about the

recipe's sweetness and flavor is provided to the user. The predefined numerical ranges are approximated in Table 5.

4 Conclusion and Future Work

A proposed mathematical formula for baking recipes was shown capable of identifying unacceptable recipes. The results also produced logical mathematical groupings of baked good recipes. Through the Random Recipe Generator, it was shown that it is possible to generate different recipes from characteristic values via ingredient constants.

The next task for both the Recipe Report Card and the Random Recipe Generator is to produce structured lists of baking recipes. Other areas of investigation include the discovery of additional ingredient constants and the continued development of the current mathematical formula to address dairy-based desserts (e.g., ice cream, cheesecake, and custards).

References

1. Corriher, S.: Bakewise: The Hows and Whys of Successful Baking. Scribner, New York (2008)
2. Ruhlman, M.: Ratio: The Simple Codes Behind the Craft of Everyday Cooking. Scribner, New York (2010)
3. Wolf, K.I., Goetze, S., Wallhoff, F.: CooCo, What Can I Cook Today? Surprise Me. Workshop Proceedings from the Twenty-Third International Conference on Case-Based Reasoning. ICCBR 2015. Columbus (2015)
4. Ohene, M.: Michael Ohene's Cookie Recipe Generator. Cooking for Engineers, http://www.cookingforengineers.com/article/309/Michael-Ohenes-Cookie-Recipe-Generator
5. Ohene, M.: Analyzing a Baking Recipe. Cooking for Engineers, http://www.cookingforengineers.com/article/280/Analyzing-a-Baking-Recipe
6. Chef Watson, http://www.ibmchefwatson.com
7. Badra F., Cordier A., Lieber J.: Opportunistic Adaptation Knowledge Discovery. In: McGinty L., Wilson D.C. (eds) Case-Based Reasoning Research and Development. ICCBR 2009. LNCS, vol. 5650. Springer, Berlin, Heidelberg (2009)
8. Gaillard E., Lieber J., Nauer E.: Improving Ingredient Substitution Using Formal Concept Analysis and Adaptation of Ingredient Quantities with Mixed Linear Optimization. Proceedings from the Twenty-Third International Conference on Case-Based Reasoning. ICCBR 2015. Columbus (2015)
9. Gaillard E., Lieber J., Nauer E., Cordier A.: How Case-Based Reasoning on e-Community Knowledge Can Be Improved Thanks to Knowledge Reliability. Case-Based Reasoning Research and Development, Sep 2014, Cork, Ireland, Ireland. 8765, 155 - 169 (2014)
10. bakeFromScratch, http://www.bakefromscratch.com/
11. Megan's Granola, http://allrecipes.com/recipe/98390/megans-granola/
12. Mary's Sugar Cookies, http://allrecipes.com/recipe/23750/marys-sugar-cookies/

	Fat Values											
	0.00-0.05		0.05-0.10			0.10-0.20			0.20-0.34			
	Egg Values											
	0-0.5	0.5-1.0	0-0.5	0.5-1.0	1.0-2.0	0-0.5	0.5-1.0	1.0-2.0	0-0.5	0.5-1.0	1.0-2.0	2.0-3.0
Moistness												
0.00-0.05												
0.05-0.10												
0.10-0.15												
0.15-0.20						streu						
0.20-0.25			gran			cook			bcook	cook		
0.25-0.30		angel	pnuss	bisco		cook	cook		bcook	cook		
0.30-0.35	bd	chal	bd	bisco	bisco	cook	cook		bcook	cook		
0.35-0.40	bd	chal	bd	bri	bisco	pie	ging		cook			
0.40-0.45	cia	chal	sco	king	king	cobb			pie			
0.45-0.50	nokn		sco	dane				bri			brown	
0.50-0.55			biscu			cobb		bri	croiss		brown	
0.55-0.60						cobb	muff		croiss			
0.60-0.65					baba		muff	kuge				
0.65-0.70							muff	b.bd			coff	lbs
0.70-0.75					sav	eng	muff	coffee			lbs	lbs
0.75-0.80							muff	coffee				lbs
0.80-0.85								cake			tea	lbs
0.85-0.90								cake			apl	lbs
0.90-0.95								cake			ct	
0.95-1.00								cake			ct	
1.00-1.05								cake			ct	
1.05-1.10												
1.10-1.15												

Table 5. Distribution of baked good characteristic values. The following abbreviations were used. (*angel*) - angel food cake; (*apl*) - apple cake; (*baba*) - *baba al rhum*; (*b.bd*) - banana bread; (*bisco*) - biscotti; (*biscu*) - biscuit; (*bd*) - bread; (*bri*) - brioche; (*chal*) - challah; (*cia*) - ciabatta; (*ct*) - carrot cake; (*coff*) - coffee cake; (*dane*) - danish; (*cobb*) - cobbler; (*ging*) - gingerbread; (*gran*) - granola; (*king*) - king cake; (*kuge*) - *kugelhof*; (*muff*) - muffin; (*nokn*) - no knead bread; (*eng*) - Old English cake; (*pie*) - pie crust; (*pnuss*) - *pfeffernuesse*; (*lbs*) - pound cake; (*sav*) -savarin; (*sco*) - scone; (*streu*) - streusel; (*tea*) - tea cake; (*bcook*) signifies *pate brisee*, butter cookies, Mexican wedding cookies, *k'ak*, *nuhood al-adhraa*, and other eggless cookies. (*cook*) signifies chocolate chip cookies, oatmeal cookies, snickerdoodles, and other cookies that contain eggs.