Recommendations for Chemists: A Case Study

Steven L. Rohall
Novartis
Cambridge, Massachusetts
steven.rohall@novartis.com

Margaret Pancost-Heidebrecht
Novartis
Cambridge, Massachusetts
margaret.pancost@novartis.com

Bill Shirley
Novartis
Emeryville, California
bill.shirley@novartis.com

Douglas Bacon
Novartis
Cambridge, Massachusetts
doug.bacon@novartis.com

Michael A. Tarselli
Novartis
Cambridge, Massachusetts
mike.tarselli@novartis.com

ABSTRACT

Large pharmaceutical companies have a wealth of reaction and chemical structure data, but face a new problem: analyzing that corpus to yield project insights and future directions. One straightforward approach would be to have a recommendation system to match drug structures with similar research endeavors across geographically- or organizationally-separated groups. We developed and deployed *Chem Recommender*, a system that suggests similar, related work to experiments that chemists have recently started. The goal of the system is to accelerate the drug discovery process by ensuring that chemists are aware of each other's work. To date, we have sent more than 8500 recommendations to over 800 medicinal chemists in our organization. The results have been positive, with several chemists reporting that the recommendations have aided their molecular syntheses.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; *Chemical and biochemical retrieval*; • **Human-centered computing** → *Collaborative and social computing systems and tools*; • **Applied computing** → *Life and medical sciences*;

KEYWORDS

medicinal chemistry; recommendations; electronic lab notebook

1 THE PROBLEM OF DRUG DISCOVERY

Drug discovery is a long and expensive process. New drugs can take 12-15 years, and cost billions of dollars, to get to market [7]. Large companies often initiate parallel projects investigating similar drug targets; it is quite possible that scientists in different locations, or even in the same lab, are doing similar work. We believe that, by alerting chemists to related work, we allow them to accomplish their goals more quickly, saving both time and money.

RecSys '18, October 2–7, 2018, Vancouver, Britsh Columbia, Canada

ACM ISBN 978-1-4503-5901-6/18/10…$15.00
https://doi.org/10.1145/3240323.3240376

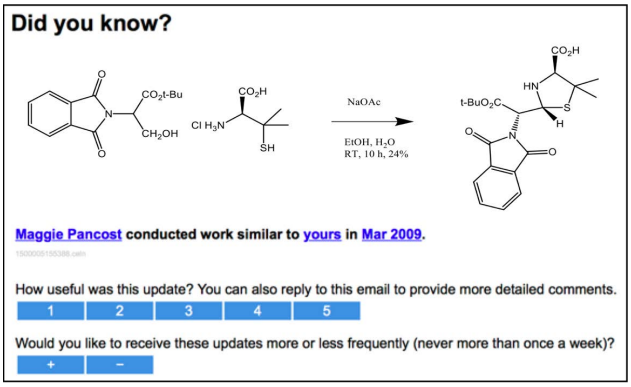

Figure 1: An example recommendation email. (Reaction shown from the synthesis of penicillin [15].)

To improve knowledge sharing across the Novartis Institutes for BioMedical Research, Novartis's drug discovery organization, we implemented *Chem Recommender*, a recommendation system that analyzes work that chemists have recently completed in their electronic lab notebook (ELN) and suggests prior, related work to them. Chem Recommender is a simple version of a case-based recommender system: a chemist's recent experiment is used as the input for a similarity search across all experiments [1, 10]. The similarity search is based upon a textual representation of the reactants and product(s) involved in a reaction. Unlike traditional case-based systems, Chem Recommender initiates the search on behalf of the user (on a daily basis) and does not provide a user interface to support a conversation with the user to refine the recommendation.

In the current implementation, the recommendations are sent via email; we would like to provide recommendations within the lab notebook directly, but are unable to modify that system's user interface. Since we launched Chem Recommender in July 2017, we've sent over 8500 recommendations to more than 800 unique chemists. At this time, only the top hit is sent to the chemist—the goal being to produce a recommendation that can be consumed quickly, without scrolling or clicking. The similarity search includes a time component that favors older experiments—when faced with a choice of several candidates, we choose the oldest experiment on the assumption that the chemist is less likely to be familiar with older work or the chemists who have performed it. We also ensure that we do not re-recommend the same scientist or experiment to the same person.

Figure 1 contains a sample recommendation email. Each recommendation includes an image of the experiment being recommended and links to the other scientist and the recommended ELN entry. It also includes a link to the ELN entry that was the source of the recommendation, providing an explanation to the recipient. Other buttons allow a one-click rating and allow the scientist to increase or decrease the frequency of the recommendations. We only send an email when we have something to recommend; being a non-interactive system, there is no pressure to generate a recommendation for the user if we have nothing useful to suggest.

2 GENERATING RECOMMENDATIONS

Since 2008, all of our medicinal chemists have used an online ELN to record their work. Every chemist is able to see any lab notebook entry; however, it is a manual process that the chemist must perform outside of the laboratory. As bench chemists may have 3-10 reactions in process on any given day, they will not search conditions or substructures except on complex substrates or problematic reactions.

The Chem Recommender has indexed over 1.5 million experiments containing nearly 6 million molecules from the chemistry ELN. Our index takes advantage of the isomeric SMILES representation of the compounds to enable similarity searching [17]. Isomeric SMILES is a textual representation of a molecule that is both canonical—the same for all occurrences of the molecule—and includes 3D molecular information [18]. SMILES strings are used as the basis for a term frequency-inverse document frequency (TFIDF) index which can be used to determine the similarity of molecules [13]. Every reactant and product molecule is separately indexed, pointing to its reaction.

To create our index, we generate n-grams of length 4, called LINGOS, from the SMILES strings [5, 11, 16]. These n-grams serve as "keywords" in a "document" that is the molecule's SMILES string. We use Elasticsearch to create the TFIDF index [4]. n-grams that are common will have a high document frequency and will not be differentiating whereas n-grams that are more distinct will have a low document frequency. To find similar molecules, we process the SMILES string of the search molecule into n-grams in the same manner and then search for reactions that have products with overlapping n-grams. Although this technique is simple, it performs as well as geometric- or fingerprint-based similarity comparisons.

To achieve useful recommendations, we look at the entire reaction in an experiment, not just the products. Our algorithm determines that two experiments are similar if at least one product is 90%+ similar and there is 80%+ similarity in at least half of the reactants.

While the search mechanism we employ for our recommendations is consistent with the manual search capability in the ELN, it uses a different mechanism. The ELN search is based upon geometric substructures; performing a complete reaction search, while possible, is not widely used. When encountering a problematic synthesis, a chemist will typically perform a substructure search looking for reaction products that have a similar substructure and then view the reaction used to synthesize those products. It is important to note, though, that our recommendations are not intended to replace these manual searches. Rather, our intent is to help scientists become more familiar with each other's work—providing serendipity in a large, distributed organization.

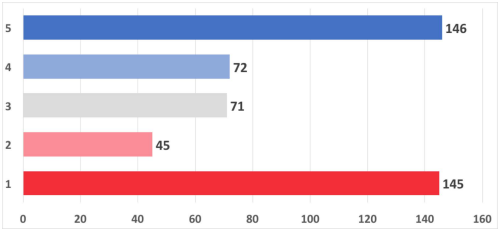

Figure 2: Rating distribution, 1 (low) – 5 (high).

3 RESULTS

Our goal is to inform chemists' work—whether or not to pursue certain reactions or target molecules. This is difficult to measure since our recommendation emails are intentionally designed to provide as much information as possible without requiring any action on the part of the receiver. It is possible that a chemist will learn something just by seeing the other chemist's name or an image of the reaction being recommended; the lack of explicit user action on the email does not indicate that the recommendation was not useful.

As a result, we have taken a multi-pronged approach to determining how well our recommendation system is performing. We monitor click-throughs, allow scientists to rate the recommendations, track email frequency changes, and record comments through both online surveys and interviews.

3.1 Click-throughs

We have instrumented the links in the recommendations so that we can record click-throughs. A typical email marketing campaign will have a click-through rate of less than 4% [8]. Ours is around 11%. Several factors contribute to this increased rate: the email is sent from an internal corporate address and the scientists were informed at the start of the project that we would be sending recommendations. However, we can also attribute some of that rate to genuine interest in the content of the recommendations. Of the messages that had a click-through, 75% had a click on the link to the recommended experiment in the ELN, 42% had a click on the link to the other scientist, and 30% had a click on the scientist's own experiment that generated the recommendation.

3.2 Ratings

As seen in the example, we also provide a simple way for users to rate a recommendation on a 1 (low) — 5 (high) scale. Six percent of the recommendations we have sent have been rated. The distribution of ratings is shown in Figure 2.

We consider this distribution a success, particularly if we add 1s and 2s (190) and the 4s and 5s (218). The positive ratings slightly outnumber the negative ratings. One problem with our rating system, though, is that, it conflates a bad recommendation that is not relevant (e.g., the recommended experiment is not similar to the chemist's experiment) with a good recommendation that is relevant but not useful (e.g., the recommended experiment is similar but

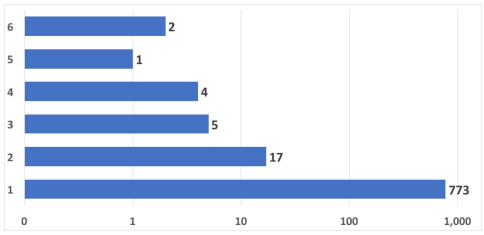

Figure 3: Number of people opting for a recommendation email every 1 – 6 weeks.

the chemist was already aware of the work). For this reason, whenever we get a negative or a positive rating, we give the chemists the opportunity to fill out a simple survey to further explain their rating. One hundred ninety-three surveys have been completed to date. In the survey, we ask if the recommendation was similar to the work they were doing, whether or not the experiment or the product is of interest, and whether the scientist already knew about the recommended work—multiple choices are permitted.

The top reason, 42% of responses, for a low rating was that the scientist already knew of the recommended work while 25% of scientists complained that the recommendation was not similar to their work. In comparison, for highly-rated recommendations, 71% of the respondents said that the recommendation was similar to their own work while 25% said that they were not already aware of the recommended work. In other words, recommendations were rated high if they were, indeed, similar; recommendations were rated low if the work was already familiar. We hypothesize two reasons for the familiarity: 1) the work is routine chemistry known to all the chemists, and 2) the recipient actually knows the other chemist and/or the experiment that was recommended (e.g., they are on the same project team).

3.3 Frequency Changes

We allow users to change the frequency of their recommendations from a maximum of once per week to a minimum of once every 6 weeks. We do not provide a mechanism to entirely opt-out of the recommendation emails. To date, we have had 251 frequency increases compared with 63 decreases.

As can be seen in Figure 3, the vast majority (773) of users have indicated a recommendation frequency of once per week. On more than one occasion we have observed chemists providing a low rating for a recommendation and then immediately choosing to increase their email frequency. Clearly, they see the potential in our system even if one recommendation did not provide them with useful information.

3.4 Comments

The scientists are always welcome to submit additional comments. The simple survey form allows them to type a free-form comment; some were also received by email. After analyzing the comments, we grouped them into the following categories. Representative comments from each of the categories are included.

3.4.1 Useful Recommendations. These comments validate our work as they are indicative of scientists who were able to make progress on their own projects due to our recommendations:

- *The exact transformation on the exact substrate I wanted. I tried another method that did not work. Thanks!*
- *This was an interesting finding so I forwarded to other people in the . . . team who will work on this scaffold soon.*
- *This was actually VERY useful! As I am trying to do the same thing on something similar and failing!*

3.4.2 Recommendations with Negative Results. Our biggest surprise was the reaction to negative results. We imagined at the start of the project that informing scientists about experiments that did not succeed would help them avoid similar problems. In practice, negative results got low ratings:

- *The shown reaction did not yield any product.*
- *The reaction didn't work.*
- *This Experiment gave no Product.*

We eventually modified our algorithm to de-prioritize such experiments.

3.4.3 Other Comments. One thing we definitely noticed is that user's expectations are high, most likely set by their experiences with commercial recommenders in the consumer world, such as those encountered on Netflix, Spotify, and Amazon [12]. As a result, some comments seem as though they expect the recommendation system to be able to read minds:

- *I had a bifunctional starting material and was interested in selectivity not only conversion!*
- *It was a chiral separation (of which I submitted) not [a] reaction.*

In other cases, the recommendation may have been useful, but the chemists had already solved their problems:

- *It's exactly the reaction I seek, but I already saw it myself by searching in eln.*
- *My Rxn already worked. I don't care who [else] did it.*

In some cases, our recommendations just missed the mark, either the chemistry was simple or not relevant to the scientist:

- *This is standard conditions for a nosyl deprotection, nothing novel or interesting here.*
- *No value in such . . . information.*

In other cases, our recommendations were good, but recommended people who were already working closely together. In one case, a chemist was performing separations and another was purifying those separations—of course they were familiar with each other's work!

- *. . . is the person who gave me this work to do and advised me to look at his experiments, he is my direct colleague and sits just next to me.*
- *We did this experiment together on the same project*

However, we also have evidence that people in the same organization on larger projects are not always familiar with their colleagues' work, so we do not use organization or project data to limit our recommendations.

4 PRIOR WORK

Savage, *et al.* developed a system for recommending candidate reactants for the synthesis of a given product [14]. They approach the problem as a link-prediction problem over a network of chemical

reactions (reactants linking to products) described in the patent literature. Their system is intended to aid the chemist in designing an experiment to synthesize a proposed molecule, but it has not yet been deployed. Rather, they have concentrated on developing and comparing algorithms that take advantage of their chemical network to improve on a similarity-based approach. Although they are using recommender system algorithms in a very interesting way, they are not providing recommendations within the context of chemists' work, as we are.

Hall, *et al.* describe using a graph database to efficiently search for close analogs of a given molecule [6]. While their data structure could be used for experiment recommendations, their main concern was in improving search for an individual chemist who would like to get a better sense of the chemical space around a particular molecule or fragment.

Corkery proposed, but did not implement, applying standard collaborative filtering techniques to the task of ordering chemical compounds from a supplier, the idea being that additional molecules would be suggested based upon the contents of a chemist's shopping cart [2]. This is no different than existing shopping recommenders and may actually be more difficult to implement since, without knowledge of what the chemist is planning to do with a given molecule, it would be difficult to recommend additional molecules out of the hundreds or thousands that could be suggested that would be useful to the chemist.

5 DISCUSSION

5.1 Collaborative Filtering Algorithms

We initially considered collaborative filtering approaches for our system, but we lack the rating data upon which these systems rely [9]. Our chemists are synthesizing molecules as part of their jobs, so it would be a stretch to say that they "like" a molecule in the same way as buying a product or rating a movie. Indeed, the synthesis of a particular compound may require several intermediate steps; in many cases, these intermediate steps are routine chemistry—necessary, but not particularly interesting. Similarly, a molecule that is interesting in one experiment may be mundane in another. And, even for a large pharmaceutical company, our user population is relatively small—about 800 medicinal chemists in the research organization. As a result, we do not have enough data to develop cohorts, either of our users or our items (molecules or experiments), as you would normally use with collaborative filtering algorithms.

5.2 Molecule and Experiment Attributes

We have relied on molecular structure for our similarity search. Molecules have many attributes beyond their structure, e.g., molecular weight and melting point. However, these are not useful for similarity search and recommendation. That is, molecular weight is an interesting attribute and it may affect the experiment that a chemist performs, but chemists do not search for molecules that weigh 342.297 g/mol. Chemists are interested in a molecule's function. In this way, molecule attributes are not like other product attributes that could be used for recommendation (e.g., *red* cars, *classical* music, *mystery* novels). It is for this reason that traditional content-based recommendation algorithms are not applicable for our use case. Chemists are interested in similar experiments, but similarity cannot be determined by looking at the attributes of the molecules involved in the experiment nor in attributes of the experiment itself (e.g., yield, temperature, time).

5.3 Diversity in Recommendations

A typical complaint of a case-based recommendation approach is that it does not do a good job with diversity or novelty. These are not a concern in our environment, though. In fact, in early testing, when we recommended some inscrutable experiments to chemists (quite by accident, usually), the feedback was universally negative. Scientists were most interested in information that could help them today or tomorrow in the lab, not in learning about some unrelated experiment.

6 FUTURE WORK

First, we are always looking for improvements to our similarity algorithm. For example, if an experiment has only one starting material and two or more identical products, then it is most likely a purification reaction—these are not interesting to recommend. There are also functional classifications of molecules such as MACCS (Molecular ACCess System) keys [3]. These keys describe molecules by their features (e.g., "is an alcohol", "contains a halogen")—features which imply a particular chemical reactivity. These features may be a way of more accurately separating "interesting" reactions from "uninteresting" ones.

Second, we are looking to expand our data sources beyond the ELN. One comment we received was that, although the recommendation the chemist received was good, they would have preferred to know about it before they had completed the experiment. Our organization has an electronic molecule "white board" where chemists post molecules they are considering for synthesis. We have begun looking at these data as a source of our recommendations.

Third, we have only begun a demographic breakdown of the users of our recommendations. We think it would be interesting to look at those who have rated recommendations and break them apart by level, tenure, time in industry, location, organization, etc. These data could help us target future recommendations to particular subpopulations for maximum impact.

7 CONCLUSION

We have presented Chem Recommender, a recommendation system to support the work of our medicinal chemists. Our goal has not been to develop new recommendation algorithms; rather, our goal is to accelerate the drug discovery process. Chem Recommender does this by analyzing the ELN entries for work chemists are currently doing, and recommending similar experiments previously completed. While Chem Recommender is an ongoing project, we have evidence that some of our recommendations have already helped the chemists. Evaluation is difficult as our recommendations are designed to provide as much information to the chemist as possible without requiring them to do any additional work. The data we do have indicate that eliminating "uninteresting" experiments from the recommendations would be useful future work.

REFERENCES

[1] D. Bridge, M.H. Göker, L. McGinty, and B. Smyth. 2005. Case-based Recommender Systems. *Knowledge Engineering Review* 20, 3 (Sept. 2005), 315–320. https://doi.org/10.1017/S0269888906000567

[2] J. Corkery. 2011. Recommendation System for Compound Selection. (Feb. 2011). https://www.eyesopen.com/blog/2011/02/28/recommendation-system-for-compound-selection

[3] J.L. Durant, B.A. Leland, D.R. Henry, and J.G. Nourse. 2002. Reoptimization of MDL Keys for Use in Drug Discovery. *Journal of Chemical Information and Computer Sciences* 42, 6 (2002), 1273–1280. https://doi.org/10.1021/ci010132r

[4] Elastic. 2018. Elasticsearch. (2018). https://www.elastic.co/products/elasticsearch

[5] J.A. Grant, J. A. Haigh, B.T. Pickup, A. Nicholls, and R.A. Sayle. 2006. Lingos, Finite State Machines, and Fast Similarity Searching. *Journal of Chemical Information and Modeling* 46, 5 (2006), 1912–1918. https://doi.org/10.1021/ci6002152

[6] R.J. Hall, C.W. Murray, and M.L. Verdonk. 2017. The Fragment Network: A Chemistry Recommendation Engine Built Using a Graph Database. *Journal of Medicinal Chemistry* 60, 14 (2017), 6440–6450. https://doi.org/10.1021/acs.jmedchem.7b00809

[7] J.P. Hughes, S. Rees, S.B. Kalindjian, and K.L. Philpott. 2011. Principles of early drug discovery. *British Journal of Pharmacology* 162, 6 (March 2011), 1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

[8] IBM. 2016. 2016 Email Marketing Metrics Benchmark Study. (2016). https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=UVL12406USEN

[9] J. A. Konstan and J. Riedl. 2012. Recommended for you. *IEEE Spectrum* 49, 10 (Oct. 2012), 54–61. https://doi.org/10.1109/MSPEC.2012.6309257

[10] F. Lorenzi and F. Ricci. 2005. Case-Based Recommender Systems: A Unifying View. In *Proceedings of the 2003 International Conference on Intelligent Techniques for Web Personalization*, B. Mobasher and S.S. Anand (Eds.). Springer-Verlag, 89–113. https://doi.org/10.1007/11577935_5

[11] H. Öztürk, E. Ozkirimli, and A. Özgür. 2016. A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. *BMC Bioinformatics* 17, 128 (March 2016). https://doi.org/10.1186/s12859-016-0977-x

[12] A. Purakayastha. 2017. The Consumerization Of Enterprise Technology. (Dec. 2017). https://www.forbes.com/sites/forbestechcouncil/2017/12/04/the-consumerization-of-enterprise-technology

[13] G. Salton and M.J. McGill. 1986. *Introduction to modern information retrieval*. McGraw-Hill, Inc.

[14] J. Savage, A. Kishimoto, B. Buesser, E. Diaz-Aviles, and C. Alzate. 2017. Chemical Reactant Recommendation Using a Network of Organic Chemistry. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 210–214. https://doi.org/10.1145/3109859.3109895

[15] J. C. Sheehan and K. R. Henery-Logan. 1957. The Total Synthesis of Penicillin V. *Journal of the American Chemical Society* 79, 5 (1957), 1262–1263. https://doi.org/10.1021/ja01562a063

[16] D. Vidal, M. Thormann, and M. Pons. 2005. LINGO, an Efficient Holographic Text Based Method To Calculate Biophysical Properties and Intermolecular Similarities. *Journal of Chemical Information and Modeling* 45, 2 (2005), 386–393. https://doi.org/10.1021/ci0496797

[17] D. Weininger. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* 28, 1 (1988), 31–36. https://doi.org/10.1021/ci00057a005

[18] D. Weininger, A. Weininger, and J.L. Weininger. 1989. SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences* 29, 2 (1989), 97–101. https://doi.org/10.1021/ci00062a008