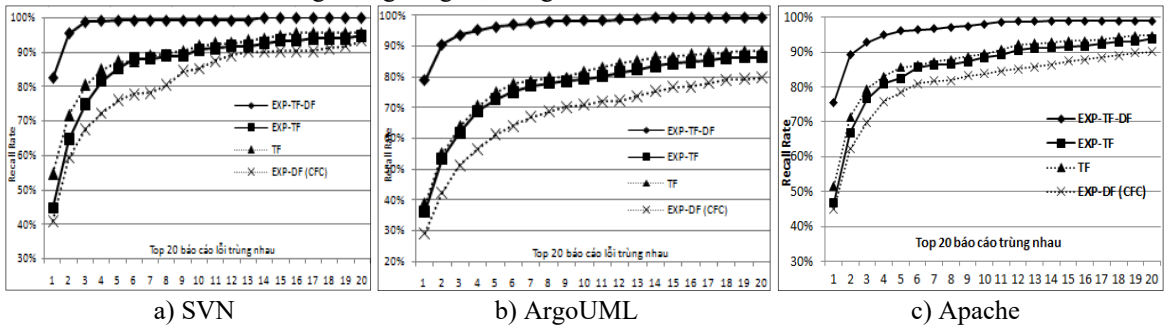


4.2 Những nhân tố tác động đến phương pháp dò tìm

Trong phần này muốn thảo luận những nhân tố ảnh hưởng đến phương pháp được giới thiệu. Thứ nhất là công thức được chọn cho việc tính trọng lượng đặc điểm lớp. Thứ hai và thứ ba liên quan đến việc xác định giá trị tốt nhất cho các tham số b , d , và h . Cuối cùng là công thức tính sự giống nhau giữa hai báo cáo lỗi sử dụng denormalized cosine.

4.2.1 Trọng lượng đặc điểm lớp

Trong phần trước đã giới thiệu bốn công thức khác nhau cho việc tính trọng lượng từng từ, trong



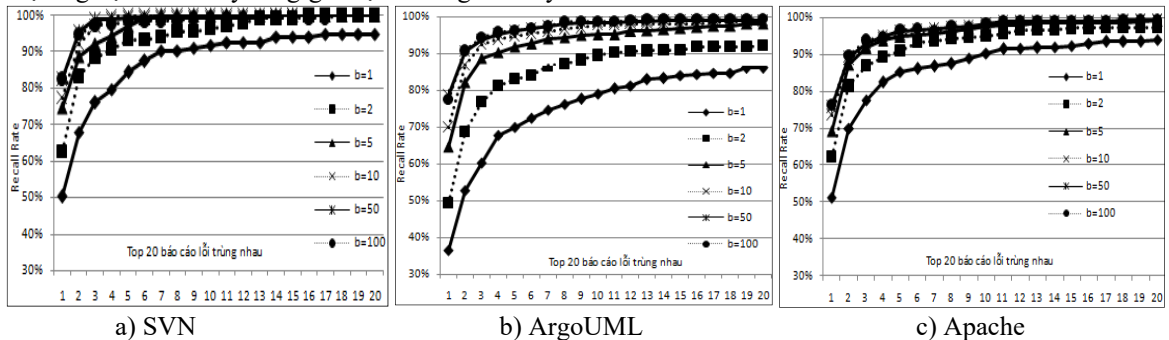
Hình 7: So sánh bốn dạng công thức tính trọng lượng lớp Inner

4.2.2 Tham số b

Tham số b đóng vai trò quan trọng trong EXP-TF-DF, do đó việc thực nghiệm để tìm ra giá trị tốt nhất cho tham số b là cần thiết để tìm ra giá trị b tốt nhất trong phương pháp của chúng tôi. Kết quả thực nghiệm cho thấy rằng giá trị b không có thay

những báo cáo lỗi dựa vào lớp bao gồm: EXP-DF, TF, EXP-TF, và EXP-TF-DF. Khi tiến hành thực nghiệm để tìm ra công thức tốt nhất cho việc tính trọng lượng đặc điểm lớp, qua quan sát kết quả thực nghiệm cho thấy rằng EXP-TF-DF cho kết quả tốt nhất trong bốn công thức. Lý do EXP-TF-DF cho kết quả tốt nhất có thể giải thích là do hỗ trợ nhiều cho thông tin từ dựa vào lớp, điều này cũng giải thích lý do CFC cho kết quả không tốt bởi nó đã không xem xét sự tác động những từ thường xuyên xuất hiện trong báo cáo lỗi dựa vào lớp. Hình 7 cho thấy sự vượt trội của phương pháp EXP-TF-DF.

đổi nhiều khi b lớn hơn 50, trừ dự án SVN có tác động nhỏ nhưng không đáng kể. Do đó, trong phương pháp này đã sử dụng $b=50$ cho các thực nghiệm còn lại. Tuy nhiên, giá trị b có thể sẽ có thay đổi tùy dữ liệu thực nghiệm. Hình 8 cho thấy kết quả thực nghiệm đối với tham số b .



Hình 8: Tìm giá trị tốt nhất cho tham số b

4.2.3 Tham số d và h

Tham số d và h cũng có ảnh hưởng trực tiếp đến EXP-TF-DF, do đó việc xác định giá trị tốt nhất cho d và h cũng góp phần quan trọng trong phương pháp được giới thiệu. Do đó, việc thực nghiệm cũng được tiến hành để thấy sự ảnh hưởng của d và h trong EXP-TF-DF, với $0.3 \leq d \leq 0.8$

và $1.5 \leq h \leq 20$. Do có nhiều sự kết hợp giá trị giữa d và h , trong bài báo này chỉ trình bày một vài trường hợp minh họa chính để tìm ra giá trị tốt nhất cho d và h . Từ kết quả thực nghiệm như trong Hình 9 đã xác định được giá trị tốt nhất cho d và h với $d=0.3$, $h=1.5$. Tuy nhiên, giá trị này có thể thay đổi tùy theo những tập dữ liệu khác nhau.